Intelligent Design Objects (IDO)

A cognitive approach for performance-based design

Michael S. Bittermann

Intelligent Design Objects

Michael S. Bittermann

Intelligent Design Objects (IDO)

A cognitive approach for performance-based design

Design is complex. This is because it involves conflicting goals that are often vague. Also, prior to the design it is generally not clear how important goals are relative to each other. And finally the amount of possible solutions is large in general. These bottlenecks are addressed in this thesis.

A novel approach for design is proposed, where computation is used to reach most suitable solutions. The approach is based on a novel concept of the objects forming a design. This concept is termed *intelligent design objects.* Such objects exhibit intelligent behavior in the sense that they approach most desirable solutions for conflicting, vague goals put forward by a designer. That is, the objects know 'themselves' what to do to satisfy the designer's goals. This is accomplished using methods from the domain of computational intelligence, as these are uniquely able to deal with the complexity of design mentioned above.

The result from the approach is that designers and decision makers have great certainty about the satisfaction of their goals and are able to concentrate on second order aspects they were not aware of prior to the execution. The approach is implemented for two applications from the domain of architecture demonstrating its effectiveness.

The thesis addresses to students, researchers and executives in the field of architecture, and other areas of design. It may be also interesting for researchers in the domain of computational intelligence, as it provides a formalism of intelligent design, and it exemplifies the use of these modern technologies in the design domain.

A doctoral dissertation

ISBN 978-90-8570-409-6

TU Delft Delft University of Technology

Intelligent Design Objects (IDO)

A cognitive approach for performance-based design

Michael S. Bittermann

Propositions

1. Although designing is practised for thousands of years it is not known precisely what it is or how it works.
2. A phenomenon appears mysterious, even magical, until we know enough about it.
3. Pareto optimality and fuzzy sets will be key concepts in a computational theory of design.
4. A human brain is so complex that modelling its cognitive functionality using deterministic means is like trying to predict the outcome of a coin toss by predicting the precise trajectory of the coin.
5. Visual perception is probabilistic, visual awareness is deterministic.
6. Natural language is a limited means of communication. This is because the meaning of words is ambiguous. Therefore the language of science is mathematics.
7. In the principle statement of modernism ‘form follows function’ the word ‘follows’ refers to a search for optimality.
8. Thinking is one of a human’s most powerful assets. Thinking that this asset is perfect to deal with the complexity of reality is the serious misleading of this asset.
9. Evolutionary computation became revolutionary with the advent of computer technology.
10. Cognition requires intelligence and not vice versa.
11. The reasons behind our activities are more complex than we are aware of. This is why we can surprise ourselves and others.
12. Soft sciences, like part of architecture, can advance as a scientific discipline saliently only with the presence of methods from hard sciences.

Deze stellingen worden verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotoren.

Intelligent Design Objects (IDO)

A cognitive approach for performance-based design

Michael S. Bittermann

Stellingen

1. Ondanks dat het ontwerpen sinds duizenden jaren gepraktiseerd wordt, is het niet precies bekend wat het is of hoe het werkt.
2. Een fenomeen lijkt mysterieus, of zelfs magisch te zijn, tot dat we er genoeg over weten.
3. *Pareto optimaliteit* en *fuzzy sets* zullen sleutelconcepten zijn in een *computational theory* van ontwerpen.
4. Het menselijk brein is zo complex, dat bij het modelleren van zijn cognitieve functies door gebruik van deterministische middelen het zo is, alsof men probeert het resultaat bij het werpen van een munt te voorspellen door het exacte vliegtraject van de munt te voorspellen.
5. Visuele waarneming is probabilistisch, visueel bewustzijn is deterministisch.
6. Natuurlijke taal is een beperkt communicatiemiddel. Dit is omdat de betekenis van woorden onduidelijk is. Daarom is mathematiek de taal van de wetenschap.
7. In de principiële stelling van het modernisme *'vorm volgt de functie'*, refereert het woord 'volgt' aan een zoektocht naar optimaliteit.
8. Het denken is een van de meest krachtige vaardigheden van de mens. Te denken dat men met deze vaardigheid perfect is toegerust om met de complexiteit van de realiteit om te gaan, is de grootste illusie van deze vaardigheid.
9. *Evolutionary computation* is revolutionair geworden door de vooruitgang van de computer technologie.
10. Cognitie heeft intelligentie nodig, en niet andersom.
11. De redenen achter onze handelingen zijn complexer dan wij bewust zijn. Daarom kunnen we onszelf en anderen verrassen.
12. Zachte wetenschappen, zoals deel van architectuur, kunnen als wetenschappelijke disciplines alleen met de aanwezigheid van methodes uit de harde wetenschappen noemenswaardige vooruitgang boeken.

Deze stellingen worden verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotoren.

Intelligent Design Objects (IDO)

A cognitive approach for performance-based design

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 6 oktober 2009 om 15.00 uur
door Michael Stefan BITTERMANN
bouwkundig ingenieur
geboren te Würzburg, Duitsland.

Dit proefschrift is goedgekeurd door de promotoren:

Prof.dr.ir. I.S. Sariyildiz
Prof. dr.ir. Ö. Ciftcioglu

Copromotor: Dr.ir. R.M.F. Stouffs

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof.dr.ir. I.S. Sariyildiz	Technische Universiteit Delft, promotor
Prof.dr.ir. Ö. Ciftcioglu	Istanbul Technical University, promotor
Dr.ir. R.M.F. Stouffs	Technische Universiteit Delft, copromotor
Prof.dr.ir. U. Kaymak	Erasmus University Rotterdam
Prof.dr.ir. J.C. Paul	Technische Universiteit Delft
Prof.ir. P.G. Luscuere	Technische Universiteit Delft
Dr.ir. L. Caldas	Technical University of Lisbon, Portugal

ISBN 978-90-8570-409-6
NUR 955

To my wife Elsbeth and our daughter Noé

Preface

This PhD thesis is written at Delft University of Technology's Architecture Faculty, at the chair of Design Informatics.

The candidate wishes to express his deep gratitude to his mentors, Professor I. Sevil Sariyildiz and Professor Özer Ciftcioglu. This work is only possible due to Professor Sariyildiz' vision and persistence that adequate use of informatics in architecture is the essential step to reach advancement in design technology, and ultimately will lead to a built environment satisfying human needs to a greater extent. She directed the research towards the crucial bottlenecks in architectural design, where the complexity of design causes undesirable results when conventional means of designing are used. These bottlenecks could be overcome, because Professor Ciftcioglu contributed his vast knowledge on advanced computational methods and his creativity, developing novel methods that are powerful enough to handle the formidable complexity of the design subject matter. Due to this extraordinary multidisciplinary effort, spanning from architectural design to computational intelligence, it was possible to gain deeper insight into the design process, developing a novel computational approach for architectural design. The approach may serve as formalism for architectural design, contributing to the maturity of the science of design. The author also wishes to thank his adjunct supervisor Dr. Rudi Stouffs for his insightful comments and especially for the discussions addressing the relations of the work with existing and on-going research in computational design.

The author thanks the examination committee members for evaluating the manuscript - Professor Luisa Caldas, Professor Uzay Kaymak, Professor Joop C. Paul, Professor Peter G. Luscuere, and Professor Patrick Teuffel. He also thanks Professor Kalyanmoy Deb for his helpful comments, in particular for pointing out relevancies to the innovization concept. He thanks the urban design firm Diederendirrix for providing information used in one of the applications of the approach developed. He also thanks his colleagues from the design informatics chair for the helpful discussions. He thanks all colleagues and students that participated in the perception experiments conducted, and also gratefully acknowledges the contributions by Mrs. Paula van den Berg's and Mrs. Sjade van Arum to the perception studies, and he thanks Wöhrmann Print Service.

The thesis primarily addresses architects, as well as students and researchers in the field of architectural design, who wish to inform themselves on the state-of-the-art use of computation in their domain that aims to alleviate the design complexity. Beyond that the thesis is relevant for decision makers in other areas, where criteria are complex and conflicting, such as strategic decision making in industry and other design domains. It may also be relevant to researchers from the field of computational intelligence, since the thesis presents a formalism of architectural design. This way the application potential of computational intelligence methods in architecture is clarified.

Intelligent Design Objects (IDO)
A cognitive approach for performance-based design

Michael S. Bittermann

4. Design generation

Summary

Intelligent Design Objects (IDO)
A cognitive approach for performance-based design

Architectural designs have far reaching impact on our lives. They satisfy our basic needs for safety, shelter and social interaction. Beyond that they address our need to live in an aesthetically appealing environment, and they form a contribution to our culture. This implies that a building should satisfy multiple criteria at the same time. It should be functional, aesthetic, energy efficient, cost effective, etc. As some criteria are conflicting, a design does not fulfil each criterion totally at the same time. Therefore we tolerate to some extend that they are not totally fulfilled. Also some requirements refer to non-technical aspects of design, such as visual perception aspects. Such requirements are difficult to grasp, i.e. they are *soft*. These issues make it problematic to assess the overall suitability of a design. It is difficult to be conscious of the softness of design aspects and their manifold interdependencies to make such an assessment with certainty. Next to that the solution options are manifold. A vast amount of potential solutions should be investigated to be sure we did not miss a most favourable one. This is especially important because a slight variation of a certain design parameter may influence many criteria at the same time, and drastically improve or worsen the design. However, our time is limited to experiment with the design parameters. These factors make architectural design a complex task. The complexity may lead to designs that do not satisfy the criteria to a desirable extent. Unexpected problems occur in late phases of a design when they are costly to correct. Even they may be discovered *after* the erection of a building. This becomes more and more severe as the pressure to use resources with care increases.

The motivation of the present thesis is to help architects to deal with the complexity of their task. In particular the goal is to provide means to reach designs that match multiple, soft criteria with increased certainty. This is accomplished by using advanced computational means that take care of several aspects of the design complexity. In particular the design objects, i.e. the physical elements that constitute a design are equipped with a *reasoning* mechanism. This way they 'know themselves' what to do to satisfy the architect's requirements. This means they exhibit intelligent behaviour. Using intelligent design objects an architect is able to focus on higher-level aspects of a design. Such aspects are whether the priorities among criteria are adequate, and whether the right criteria are applied. The objects take care of the primary criteria put forward by an architect, so that one can exercise preferences with great awareness, being certain that solutions match the criteria.

This approach is novel in particular as it is able to deal with *soft* issues in design. Existing computational methods, including the methods from the domain of classical artificial intelligence, are unable to deal with these issues. For example the novel approach enables a computer to handle demands for 'high visual privacy' or 'high functionality.' Such issues are conventionally considered to be incomputable, so that they form a bottleneck in evaluating the performance of a design. The advanced computational means used in the thesis to deal with such matters are known as *soft computing* methodologies. They are especially equipped to handle the complexity of design that is due to the softness of design criteria.

In particular visual perception related aspects are vaguely understood and difficult to treat. This is handled in the approach by using a probabilistic visual perception model. The model is employed to assess perceptual properties of spaces with precision. Based on the results subtle perceptual differences can be weighed against other criteria, ensuring that they are taken into account together commensurately. The latter process is accomplished by using a special model for performance evaluation. It takes the softness of criteria into account during evaluation. Based on the results from the evaluation an algorithm is employed that generates design alternatives aiming to satisfy the criteria maximally. The generation process yields a range of suitable solutions. These solutions match the criteria outstandingly, while each one of them has different advantages and disadvantages. Exploring the trade-offs among the solutions an architect is able to make more confident statements on the importance among criteria. This is because one can directly see the implications one's preferences have on the solutions.

During our exploration among the solutions we do not need to bear the multitude of relations among design objects in mind. This is because they have already been taken into account in the computations. This way we are able to focus on the interplay among complex criteria rather than having to ensure the satisfaction of elemental requirements at the same time. This means we are able to exercise our preferences with great awareness. As a metaphor we can say we are conducting an orchestra focusing on the higher-level aspects of the peace, in place of being busy to play every instrument at the same time.

The approach is applied in two architectural design applications. One of the design tasks was devised from a real scene for computation; the other one was an actual project. In the latter application the computational results are compared against the solutions obtained using conventional means. The applications differ with respect to scale and context. They demonstrate the effectiveness and generic character of the approach. It is noted that possible applications of the method are diverse, due the generic character of the methodologies involved. That is, different requirements and intelligent objects can be formulated following the same approach. Therefore the cognitive approach is applicable throughout the various phases of the building process. These range from initiative to demolition phase, while the method remains essentially the same.

Beyond architectural design the cognitive approach developed is relevant for decision making in complex problem domains, such as other areas of design and planning, including urban design, industrial design, and regional planning. It is emphasized that the approach is uniquely suitable when the problem includes multiple linguistic criteria, which is commonly the case in real-world applications. Next to decision makers and researchers in these domains the thesis may also be of interest to researchers from the field of computational intelligence, who are interested in the particular problems of architecture, and the application potentials of their methodologies in this domain.

Michael S. Bittermann

Samenvatting (Summary in Dutch)

Intelligente Ontwerp Objecten (IDO)

Een cognitieve aanpak voor prestatie-gebaseerd ontwerpen

Architectonische ontwerpen beïnvloeden ons leven ingrijpend. Zij vervullen onze basale behoeftes zoals veiligheid, onderdak, en sociale interactie. Bovendien zorgen zij ervoor dat onze leefomgeving esthetisch gezien aangenaam is en vormen zij een toevoeging aan onze cultuur. Dit betekent dat een gebouw tegelijkertijd aan meerdere criteria moet voldoen. Het moet functioneel zijn, esthetisch, energie-efficiënt, kosteneffectief etc. Sommige criteria zijn tegenstrijdig. Daarom vervult een ontwerp ook niet tegelijkertijd alle criteria volledig. Dit wordt tot op zekere hoogte getolereerd. Sommige eisen hebben betrekking op niet-technische aspecten van ontwerpen, zoals aspecten van visuele perceptie. Zulke eisen zijn moeilijk te verifiëren, omdat ze vaag en geen precieze beschrijving hebben zijn. We zeggen dan dat deze eisen *zacht* zijn. Tegenstrijdigheden en zachtheid van criteria maken het problematisch om te beoordelen hoe geschikt een ontwerp is. Het is moeilijk om bewust te zijn van de specifieke zachtheid van ontwerp aspecten en hun onderlinge afhankelijkheden, om zo een trefzekere beoordeling te maken. Daarnaast zijn er vele mogelijke oplossingen. We zouden een grote hoeveelheid potentiële oplossingen moeten onderzoeken, om er zeker van te zijn dat we niet een meest geschikte oplossing gemist hebben. Dit punt is belangrijk, omdat een minuscule variatie van een ontwerp parameter meerdere criteria tegelijk kan beïnvloeden en het ontwerp daardoor drastisch zou kunnen verbeteren of verslechteren. Niettemin is de tijd beperkt die we hebben, om met de ontwerp parameters te experimenteren. Deze factoren zorgen ervoor dat architectonisch ontwerpen een complexe taak is. De complexiteit kan tot ontwerpen leiden, die de criteria niet tot op de gewenste hoogte vervullen. Hierdoor kunnen in latere ontwerpfases onverwachte problemen ontstaan, die kostbare aanpassingen vereisen. op in late ontwerp fases, en dan is het duur ze te corrigeren. Zelfs kan het zijn dat problemen pas na de aflevering van het gebouw ontdekt worden. Deze kwesties worden steeds belangrijker door de toenemende eisen op het gebied van duurzaamheid, zodat we steeds zorgzamer met resources om moeten gaan.

De motivatie van deze thesis is om architecten te helpen, de complexiteit van hun taak te aanvaarden. Specifiek is het doel om middelen te scheppen, zodat hun ontwerpen meerdere, zachte criteria met verhoogde zekerheid vervullen. Dit is bereikt door inzet van geavanceerde rekenmethodes, die meerdere aspecten van de complexiteit van ontwerpen aanpakken. Specifiek worden de objecten in een ontwerp uitgerust met een manier om te *beredeneren*. Op deze wijze 'weten ze zelf' wat ze moeten doen, om de eisen van een architect te vervullen. Dit betekent dat de objecten intelligent gedrag vertonen. Een architect, die intelligente ontwerp objecten gebruikt, is in staat om op aspecten van zijn ontwerp te concentreren, die op een hoger niveau liggen. Zulke aspecten zijn bijvoorbeeld de prioriteiten tussen criteria en of de juiste criteria in overweging genomen zijn. De intelligente objecten zorgen ervoor, dat de primaire criteria van een architect vervuld zijn. Hierdoor kan hij zich concentreren op zijn secundaire voorkeuren, waarbij hij zich geen zorgen hoeft te maken dat aan zijn primaire criteria wordt voldaan.

Deze aanpak is nieuw, vooral omdat het in staat is om *zachte* kwesties in het ontwerp te betrekken. Bestaande rekenmethodes, inclusief de methodes uit het gebied van klassieke kunstmatige intelligentie, zijn niet in staat om zulke kwesties aan te pakken. De nieuwe aanpak stelt bijvoorbeeld een computer in staat om rekening te houden met eisen als *hoge visuele privacy,* of *hoge functionaliteit.* Zulke eisen worden conventioneel beschouwd als niet berekenbaar. Op deze wijze vormen ze een essentiële belemmering bij het beoordelen van de prestatie van een architectonisch ontwerp. De geavanceerde rekenmethodologiën, die in deze thesis ingezet zijn om zulke zaken aan te grijpen, zijn bekent als *soft computing* methodologiën. Ze hebben de bijzondere eigenschap om met de complexiteit door tegenstrijdige criteria en zachtheid van het ontwerp om te gaan.

Vooral aspecten, die gerelateerd zijn aan visuele perceptie, worden slecht begrepen en zijn moeilijk te behandelen. Dit wordt in deze aanpak opgelost door gebruik van een waarnemingsmodel, dat op kanstheorie gebaseerd is. Het model is ingezet, om perceptie gerelateerde eigenschappen van ruimtes met precisie te beoordelen. Gebaseerd op de resultaten kunnen subtiele perceptuele verschillen afgewogen worden tegenover andere criteria. Op deze wijze wordt gewaarborgd dat zowel perceptie-eisen als andere eisen naar verhouding in overweging genomen worden. Dit afwegingsproces is gerealiseerd door inzet van een speciaal model, waarmee de ontwerpprestatie geëvalueerd wordt. Het neemt de zachtheid van criteria in acht tijdens de evaluatie. Gebaseerd op de resultaten van de evaluatie worden door een algoritme ontwerpalternatieven gegenereerd met het doel de ontwerp criteria maximaal te vervullen. Resultaat van dit proces zijn een aantal passende oplossingen. Deze oplossingen vervullen de criteria bijzonder goed, en als we ze met elkaar vergelijken blijkt dat iedere oplossing andere voor- en nadelen heeft. Een architect, die zich een beeld van deze voor- en nadelen verschaft, is hierdoor in staat met een grotere zekerheid uitspraken te doen over het relatieve belang van de criteria. Dit komt omdat hij direct kan ervaren, welke implicaties zulke voorkeuren hebben op de oplossingen. Tijdens de verkenning van de oplossingen hoeft de ontwerper niet bewust te zijn van de grote aantal relaties tussen de ontwerp objecten. Dit komt omdat deze relaties reeds door de berekeningen verwezenlijkt zijn. Op deze wijze kunnen we ons concentreren op het samenspel van complexe criteria, in plaats van gelijktijdig de vervulling van basale eisen in het oog te moeten behouden. Dit betekent dat we onze voorkeuren objectief kunnen onderbouwen. Als metafoor kunnen we zeggen dat we een concertensemble aan het dirigeren zijn, waarbij we ons kunnen richten op het hogere niveau van het stuk, in plaats van elk instrument tegelijkertijd zelf te bespelen.

Deze aanpak is ingezet in twee architectonische toepassingen. Een ontwerp taak was ontwikkeld vanuit een bestaande situatie om de berekeningen uit te voeren, de andere was een real-world project. In de tweede toepassing zijn de resultaten van de berekening vergeleken met de oplossingen, die door conventionele manier bereikt zijn. De toepassingen verschillen ten opzichte van elkaar op schaalniveau en context. Ze demonstreren de effectiviteit en het generieke karakter van de aanpak. Mogelijke toepassingen van de methode zijn divers. Dit komt door het generieke karakter van de gebruikte methodologieën. Dit betekent dat verschillende eisen en intelligente objecten geformuleerd kunnen worden, waarbij steeds de zelfde aanpak gevolgd wordt. Daarom heeft de cognitieve aanpak toepassingsmogelijkheden tijdens alle fases van het bouwproces. Deze fases reiken van initiatief tot sloop fase, waarbij de methode in essentie hetzelfde blijft.

Naast architectonisch ontwerpen is de cognitieve aanpak relevant voor besluitvorming in gebieden met complexe problemen, zoals andere gebieden van ontwerp en planning, inclusief stedenbouw, industrieel ontwerpen en regionaal planontwerp. De cognitieve aanpak is overigens als enige in staat problemen met meerdere, linguïstische criteria aan te grijpen. Dit type problemen komt vaak voor in real-world toepassingen. Naast besluitvormers en onderzoekers in deze gebieden kan de thesis ook interessant zijn voor onderzoekers uit het gebied van *computational intelligence*, die de specifieke problemen van architectuur willen leren kennen en een beeld willen verkrijgen in hoeverre hun methodes in dit gebied toepassingen kunnen vinden.

Michael S. Bittermann

1. Introduction

1.1. Objectives of architectural design

Architectural design involves *alpha*, *beta,* and *gamma* sciences [1]. Alpha sciences deal with the subjective world of beauty and moral, as expressed by the artistic, intuitive soul. Beta sciences concern the objective world of facts and logic, represented by the rational mind. Gamma sciences consider the interests of society and culture. An architect is expected to integrate aspects from all three sciences in his or her design work making the task complex, but also unique and challenging. This diversity of required knowledge stems from the fact that architecture aims to satisfy a wide range of human needs, spanning basic and higher level needs.

1.1.1. Satisfaction of basic needs

Most built environments we experience in our life were designed. It may not always have been a trained architect; however somebody determined the shape of the place and its material properties. Such decisions have great impact.

Starting in our own homes, we note that the spatial organization influences the kind of interactions we have with our family members; the shapes, materials and colours of our interiors influence how we feel. Considering such influences it is clear that architectural design has a far-reaching impact on our lives, influencing how fulfilled and happy we are and how productive. The shape of a building, the material of its envelope, and the placement of the windows also influence how much energy is needed to heat the building. The choice of construction methods and material used affects the potential to alter the spatial structure for reuse, or to recycle its components once the 'lifetime' of the building ends.

The architecture of our working environment influences how frequently we interact with our colleagues, our degree of privacy when we are working, our creativity and our efficiency. Clearly this has an impact on our economies. The architecture of our hospitals influences the effectiveness and efficiency of the treatment, and therefore our health. Our cities are a product of architectural design, and their characteristics are of great importance to our individual and collective well-being. Architectural features of our neighbourhoods influence their safety, the kind of relations we maintain with our neighbours, and our feeling of belonging to a place and a community.

These examples indicate that architectural design has an impact reaching beyond the interests of the people inhabiting the building. Buildings make use of many precious resources we share, such as soil, energy, and clean air. These resources are becoming scarcer and scarcer. Therefore the relevance of architectural design is not restricted to the present time, but it concerns future generations as well. Currently buildings consume about 40% of the energy that is being used on the globe, and 50% of the products made are being used in buildings [2]. This entails that architects have the opportunity to significantly reduce the use of natural resources and production of waste.

1.1.2. Satisfaction of higher level needs

Architecture has an impact beyond primary human needs for shelter, safety and social interaction. In the terms used by Professor Maslow, who developed a hierarchy of human needs [3, 4], we can say that architecture responds to our 'esteem-needs' and our need for 'self-actualisation.'

Esteem needs are needs associated with the human strive for a high evaluation of themselves. The aim is self-respect and respect from others. Architectural design addresses these needs. For example a residence may have a spatial structure or express certain aesthetics that resonates with the user's personal preferences. This lets the inhabitant feel confirmed regarding his or her values, and enables to share the values with other people. It is no coincidence that the piano nobile of renaissance residences is located on the first floor and has larger windows compared to the other floors in the building. The reason was to provide finer views and to avoid the noise and dust of the streets. Another example of esteem satisfaction is the Seagram building in Manhattan by Mies van der Rohe. This building manifests the generous attitude of the client providing a privately owned *public* square on the lot of the building. This was an exceptional situation in the city, making the building very popular, and earning the respect from the society. An example for a simpler form of esteem satisfaction can be found in San Giminiano in Italy. Here different families aimed to have the highest tower of the town in order to gain the respect from the community. Other examples are the recent high-rise buildings in Dubai, where impressive building-shapes reflect the sense for beauty and wealth of their clients as well as the state of the art in construction technological development.

Even if the esteem needs are fulfilled, according to Maslow one is facing another source of restlessness. It is the urge of a human being to develop his or her unique abilities and interests in order to experience a harmony with his or her environment. Maslow exemplifies this saying "a musician must make music, an artist must paint, a poet must write, if he is to be ultimately happy." The role of architecture in this respect is to offer a possibility for *transformation* to a consumer of architecture to reach such an experience of harmony. As human beings differ greatly regarding what type of experience will lead them to self-actualization, there is no general recipe to follow to let a building be effective on this level. However certain buildings appeal to many people in this respect, e.g. Mies van der Rohe's Barcelona pavilion and his residences, the cathedrals of France, the landscape architecture of the Katsura in Japan, Niemeyer's Cathedral of Brasilia and Anger's Matramandir in Auroville. These buildings apparently do not merely appeal to us on an esteem level. Their experience stimulates the awareness and expression of oneself in relation to environment in an unusual manner, experiencing a harmony with the environment. In this sense we can speak of architecture as a work of art.

Beyond its significance for the satisfaction of the needs of an individual, architecture is a part of our culture [5]. This means architecture is intertwined with social, economic and ecological aspects, environmental aspects, as well as technological and historical developments. Buildings reflect the state of technological development, such as Paxton's Crystal Palace, Uzorn's Sydney Opera House, or Roger's Centre Pompidou. They may reflect the attitude of a society regarding philosophical questions, such as the designs proposed during deconstructivism or during the post-modern movement. Architecture may reflect and even enforce a certain political attitude, as it is seen for example in the buildings by Speer or the Gruppo 7 group.

Although the satisfaction of esteem-needs, needs for self-actualization and for cultural value is not addressed by all architectural designs to the same extent, these aspects distinguish architectural design from other engineering design activities [6]. Soft aspects, such as visual perception aspects, play a major role accomplishing this level of sophistication. For example the buildings by Calatrava, Zumthor or Ando have perceptual properties that appeal to peoples' emotions, indicating that the architects have a high sensitivity to perceptual properties of space and form, and its effect on people.

Satisfaction of perception-rooted needs forms the highest level of sophistication of design, and this accomplishment is deeply involved with what can be best described as Zeitgeist. A building that was designed in line with the intellectual, cultural, ethic and political climate of its era one hundred years ago may not be appreciated the same way today. The conditions of our time inevitably influence a design and how the resulting building will be appreciated.

This dependence of architectural design on the specific cultural, political and other circumstances is valid not only for 'soft' values, such as perceptual or life-style preferences for instance. They also count for so called 'hard' aspects, such as sustainability and costs. As an example we may consider the change in architectural designs when energy prices are rising. Building shapes become more monolithic aiming to reduce surface radiation, glass facades become frameless to reduce transmission loss, and the use of natural ventilation becomes a demanded feature. However, other factors are playing a role on the decision how much to invest into a less energy consuming building. Predictions on the development of energy prices, other economic trends, climate trends, availability of materials, recyclability, etc. make design decisions difficult. That is, it is problematic to find most suitable compromises that fit the intended purpose of a project.

Concerning the satisfaction of esteem needs and needs for self-actualization a particular trend appearing in the last decades should be noted. Namely, due to the present scarcity of natural resources, satisfaction of these higher-level human needs becomes relatively less urgent compared to the satisfaction of primary human needs. Noticing this trend it appears significant to be able to deal in a measured way with objectives that concern both primary and secondary human needs. This means ways should be found, so that e.g. perception related preferences of users are weighed against sustainability concerns on a common ground. This is desirable, so that in the future, when scarcity is more severe, we are prepared to find solutions, where we sacrifice as little as we can afford regarding the satisfaction of higher-level human needs.

1.2. Complexity of architectural design

1.2.1. On the sources of the complexity

Seeing the diverse and far-reaching implications of design decisions on human lives, it is clear that if we advance architectural design, we have the opportunity to increase the quality of life in a comprehensive sense. Therefore it is a relevant objective for society to develop means to advance architectural design. This advancement requires to gain insight into the design process, and to develop design technology that helps architects to design buildings that match the needs of their clients, users, and society.

The original motivation behind architectural design is the desire to predict and control the consequences of modifying our environment by erecting a building. The consequences concern the satisfaction of the diverse goals pursued with the building. The better we are able to predict such consequences, the better we are equipped to make decisions about our built environment. Placing bricks on top of each other without a plan telling what the functionality of the construction element is in relation with other building components is clearly problematic. Such an activity is more prone to yielding an undesirable final result, than if one considers these issues in advance. The awareness of the implications one's choices have on the goals pursued is necessary to prevent undesirable results. The consideration process is conventionally facilitated by means of representational aids, ranging from the hand-sketch to computer-aided drawing, parametric design, and building information modelling. However, these tools do not support the design process significantly, since they merely store information and do not process this information. Therefore they do not help us to take the complex and complicated decisions we have to take with greater confidence compared to the conventional case. So far the existing design technology is not helping to increase the certainty that design decisions are suitable.

1.2.2. Source 1: Diversity of criteria

The difficulty of taking design decisions confidently lies in the fact that architectural designs are expected to satisfy a number of goals at the same time. The goals range from the satisfaction of basic human needs to esteem needs and needs for self-realization. This means a building should be functional, sustainable, provide a comfortable indoor climate, be cost effective, look appealing, etc. When we set out to design and draw the first lines defining the rough shape of a building, we already determine much of the buildings quality in many respects. This includes costs, functionality, and aesthetic appeal. The way in which each of these aspects is influenced by the shape we drew is usually vaguely known. This makes it problematic to determine how satisfactory a design is with respect to the criteria we impose on it.

The diversity of criteria requires that a designer has a broad range of knowledge about them. Therefore, usually architectural design is a cooperative effort involving a team of experts that contribute knowledge from diverse fields. To assess the suitability of a design it is necessary to consider the influence of the building shape, orientation, and surface material on its energy consumption; the stress in the structural members under different loading conditions, and the influence of a building's shape on wind speed in its perimeter, etc. Beyond that designers also need to know what effect an open plan structure has on functionality and social interactions etc., the influence of ceiling height on the perception of spatial intimacy, and the effect of a selected material on the ambience of the place. The simultaneous consideration of these aspects makes the task of decision makers in architectural design challenging.

A major problem resulting from the diversity of criteria is to deal at the same time with the function-related, technical aspects of a design together with the psychological, 'soft' aspects of it. It is challenging to consider soft aspects, e.g. demands for visual openness, together with e.g. needs for low energy consumption. Such different requirements may be both influenced by the same properties of a design, for instance the type of glass façade used. Experts governing such different aspects of a design have difficulty to make their reasoning explicit, to 'speak the same language' as it were. Therefore an overall evaluation of a design, where the diverse aspects are taken into account at the same time, is generally a problematic issue. The different nature of criteria makes it difficult to compare suitability of alternative designs on a common ground. The complication may be regarded as a communication problem in the first place. However this understanding should be refined, noting that the problem in the communication is not so much a matter of not being able to bring *information* across, yet it is the problem of sharing ones *knowledge*.

As a simple example let us consider a communication between an architect and a structural engineer. If an architect prefers the space to be *visually open* this has implications on the positioning and diameters of the columns in this space. The same parameters also influence the structural performance of the columns and the ceiling. In order to determine the most adequate solution for the columns it is required that the knowledge both experts have is shared. This means the architect lets the engineer understand what he means by visual openness, and how the column parameters influence this feature. In the same way the engineer puts forward his preferences regarding the structural aspects, such as structural safety, and manufacturing costs. Expressing this knowledge with minimal ambiguity is a challenging issue. Therefore it is problematic to evaluate the total suitability of different solutions regarding both aspects together.

1.2.3. Source 2: Softness of criteria

From this example we note another complication resulting from the diversity of criteria. This is the fact that some criteria are conflicting. For example a building should be inexpensive, yet provide a high level of comfort at the same time. Conflicting means that fulfilment of one requirement reduces the fulfilment of at least one other requirement. Therefore conflicting requirements cannot be simultaneously fulfilled totally, i.e. their satisfaction is partial. We can speak of a *degree* of their satisfaction.

Due to partial satisfaction, a design requirement is generally not sharply defined, but it involves some tolerance for imprecision. This makes the concept of requirement a *fuzzy* concept. The tolerance refers that we are ready to accept some deviation of a design variable from its ideal value, when, in return for this sacrifice, we have some gain on another aspect. For example an office may be required to have a floor area of ideally $20m^2$, but it would also be acceptable if it had only 17 m^2. However, in the latter case the degree of satisfaction on this issue is clearly somewhat less compared to the former case, i.e. the satisfaction in case of $17m^2$ is less than 100%.

Due to the conflicting nature of criteria and their fuzzy character it is not known precisely how a modification of a given design changes the overall suitability of the design. For example if we consider a simple design modification, say changing the location of the entrance, or choosing a different construction material for the columns of a building, it is merely vaguely known what the implications of such modifications are with respect to the total suitability of the design. Is the design getting better or worse? And how much better or worse is it getting? When a design object is modified the modification usually affects the performance of the design regarding multiple aspects, and generally these aspects have soft criteria associated with them. Therefore it is difficult for a decision maker to be aware of the implications of a design action for the overall 'goodness' of the design.

1.2.4. Source 3: Multitude of solutions options

Next to the large amount of criteria and their associated softness, another aspect makes design even more involved. This aspect is the diversification of solution options. During the industrial age the technological means to build multiplied. New building materials and building products emerged, such as composites and permeable constructions. Also manufacturing and erection technology advanced, involving high degree of prefabrication. This means the designer has a large amount of options to consider in combination, and each of the options has their unique influence with respect to the satisfaction of the design requirements.

As both, the options to realize solutions become more diverse and the requirements imposed to a building, it is clear that design is a complex task. In general terms complexity can be understood as the quality of an entity consisting of a large amount of different elements that are related in different ways, making it difficult to analyze the entity. Architectural design involves many elements: multiple criteria, multiple solution options, and multiple stakeholders and experts aiming for the solution. Criteria are related with the parameters of a design in intricate ways. With this understanding architectural design is clearly complex.

This complexity makes design a difficult task. Namely due to the softness of criteria and the multitude of solution options it is difficult to take decisions with the confidence that we did not miss a more favourable option. Due to time limitations, we are usually unable to investigate a large number of alternative solutions in order to be certain a solution we prefer is actually favourable according to the design criteria. Due to the diversity we are forced to follow 'trodden paths' to some extend to arrive at solutions. Traditionally we deal with the complexity of design tasks by analyzing the problem into parts, i.e. segmenting the solution into components, for example first dealing with the general shape of the building and then deciding about the details afterwards [7, 8]. However we know from experience that when we fix a certain property of a design at a certain moment during the process, without taking the full range of relevant aspects into account, afterwards this decision may turn out to be inferior. This is because the details of a design have an influence on the performance of the whole and vice versa. Due to the large amount of dependencies among design entities, implications of our initial decisions may not be fully overseen. Therefore, later on aspects become apparent that have an influence on the previous considerations, and it may turn out the earlier decisions were unsuitable. In such a case a design is obtained that does not match the intended purpose to the desired extent. In late phases of a design process problems occur that cannot be resolved easily anymore due to premature previous commitment. This problem is indicated by the large amount of failure costs the building industry produces. Deficiencies in the design are largely responsible for this. That is, certain aspects were not taken into account duly enough, or thoroughly enough to ensure delivery of a flawless building.

1.3. Existing treatment of the complexity and its limitations

1.3.5. From traditional means to computational means

Seeing the complexity of architectural design, it is natural that during the many centuries of architectural practise we have been looking for means to expand our ability to consider alternatives. This is motivated by the aim to better consider the implications of our ideas and look for desirable solutions. Traditional means include sketching, drawing, creating physical 3d-mock ups, and 3-D mock-ups in the computer. Also building information models (BIM) belong into this tradition, while they expand the amount and types of information represented. Representing a design allows contemplating its implications. This would be difficult to manage exclusively by means of thinking, as the complexity of design makes our considerations elusive and ephemeral. Having a representation we have more time to consider our design's problematic and advantageous features in detail, change our viewpoint, both literally and metaphorically. However everything comes at a price, and so this is the case when we use representation to aid our decision making process. Namely when we represent the solution we manifest a *specific* relation among the design objects. It is difficult to integrate the reasoning leading us to this particular solution into our representation. Therefore, when we change our criteria slightly we are unable to deduct from the drawing how to modify it in response to the new criteria. Also computer aided drawing does not facilitate this.

In face of the diversity of design requirements, solution options, and their interrelations it is hypothesised that computation can play a more significant role in design; a role that goes beyond representing a design solution. Computation should enable designers to take design decisions with confidence. The questions we answer during design should be made easier to answer using computation. Such questions are which spatial organisation to take, where to place the columns, which materials to choose etc. Representations have no active role in answering these questions, i.e. they do not reduce the complexity of a design problem but they function merely as *storage* of facts. The amount of information in design is very large - the amount of criteria, the amount of solution alternatives etc. Therefore a designer's original problem, namely to reach a solution that satisfies the criteria maximally, is not becoming significantly less severe by means of representing a solution.

Computers have a unique ability with respect to retaining information in memory and performing fast and accurate information processing. Human designers have quite different abilities. An architect usually has experienced many spaces during his or her career and accumulated a vast and unique experience and knowledge. Based on this experience he/she is able to assess the particularity of a project. He is able to determine which criteria are of particular significance, and which solution direction is viable to pursue. The aim of computational design is to make use of the benefits of computation in order to allow us to better handle the complexity of designs. This way we should be enabled to exploit the creative strength we possess to a greater extent. This means we should aim to establish a symbiotic partnership between computer and designer. In this partnership a designer can benefit from computations by using them to reduce his or her cognitive load. In other words the amount of considerations the designer is required to keep in mind is reduced. This way the practical limitations, such as time and memory [9, 10], are alleviated and he/she can arrive at results with a clarity and certainty he/she could not obtain otherwise. A premise regarding the accomplishment of this aim is that a computer should be equipped with *reasoning* capabilities, so that a designer can use computation to explore a large amount of design options in a goal oriented manner.

1.3.6. Artificial intelligence methods

Based on this premise, existing research from the 1970s until today attempted to equip computers with the ability to substantially aid during design [11-14]. This means to make use of computation in the decision making process as opposed to merely using it for information representation. However, apparently the complexity of design prevented conclusive results up till now.

The main difficulty lies in developing appropriate computational methods that are able to handle the complexity of architectural design. In particular the difficulty concerns dealing with the *soft* issues of design. That is, how to let a computer deal with the *linguistic* nature of design requirements, such as 'high visual privacy' or 'high functionality', how to handle conflicts among criteria, and to deal with the multitude of solution options?

The methods used in the existing computational design approaches stem from the domain of classical artificial intelligence (AI). These methods appeared to be promising; so many efforts were put into their application for design enhancement until today. They attempt to imitate the human reasoning process by employing *rules* as concise expressions of knowledge. Systems based on these methods became known as expert systems or decision support systems. They are successfully applied for automation in a number of fields, such as accounting, medicine, financial services, etc. They were expected to be also applicable for architectural design. However, using explicit rules to arrive at solutions is an approach that is limited in several respects, as follows.

A primary limitation of classical AI methods is that they do not allow dealing with conflicting objectives. On one hand a building may be required to be spacious; on the other hand it should be inexpensive. A rule-based system has no means to handle such a conflict. It is based on formalizing solutions for either criterion, but not for the case when both apply at the same time. The source of the difficulty these methods have with conflicts is their inability to treat requirements that are satisfied to a *degree*. Namely basis for the classic AI methods is formed by propositional or predicate logic. This logic exclusively treats conditions that are either completely fulfilled or not at all. Architectural design involves conflicting requirements. Therefore some requirements are generally partially fulfilled. As an example let us consider the conflicting nature of a requirement for low energy use of a building, while the costs for this building should be low at the same time. Increase of the satisfaction of one requirement is at the cost of a decrease for the other one. Therefore, generally either requirement is satisfied to some degree, i.e. the level of satisfaction is situated somewhere between being completely satisfied and not satisfied at all. Classical AI methods do not handle this, so that their applicability for design is severely limited.

Another limitation of classic AI methods is their rigid structure. This does not permit to adapt them easily to particular requirements. Design criteria and viable solution options depend on specific economic, cultural, personal circumstances. When we say we want a *good* building, it can mean quite different things to different people and in different situations. Clearly the ‘goodness’ of a design is a complex attribute consisting of many criteria and sub-criteria, and its meaning depends on the circumstance. In general a building should be functional, provide a pleasant visual experience, and consume little energy, for example. But then again what is meant by *functional* e.g. is imprecise and subject to clarification from case to case. Therefore approaches such as [15], that are based on rules aiming to identify general solutions to design problems, have a limited validity. This is because general rules are unable to reflect the circumstances of a certain economic, personal, etc. situation. Next to this, usually the relative importance among the criteria is uncertain initially. Therefore criteria are usually changing during the design process [7]. This makes it difficult to establish generally valid design solutions. To take the specific conditions of a context into account makes it necessary to adapt the rule base of the expert system. However, an excessive amount of rules are required to cover common design circumstances. Therefore adaptation of the rule base is unfeasible in architectural design.

In summary one notes that the rigid reasoning mechanism of the classical AI approach limits the applicability of the approach severely. The classical AI methods are unable to deal with the multiple, conflicting nature of design criteria, involving requirements that are imprecisely defined and partially satisfied. Examples of such requirements are requirements for visual perception properties of spaces, which are common in architecture. These characteristic sources of design complexity can be summarized as *soft issues*. Due to the shortcomings of the classical AI methods to deal with the soft issues, application of the methods in complex problems was severely questioned in the last decades [16, 17]. Due to the shortcoming to deal with the soft issues of design, the application of the classical AI methods for architectural design must be considered a failure. There is a gap in our scientific knowledge on how to help architects to handle the complexity of their task.

1.4. Research questions and requirements for an approach

1.4.1. Research objective and question

The research is motivated by the awareness of the far-reaching implications that the decisions we take about our built environment have, and what difficulties are associated with taking them. From the previous sections it is emphasized that the design process is a highly complex process, where multiple, conflicting requirements should be satisfied at the same time. Some of the requirements are ‘hard’, such as building regulations and site limitations. Others are 'soft', i.e. they are linguistic, imprecise, or uncertain, such as '*high functionality'* or requirements related to visual perception aspects. It is also emphasized that up till now science did not provide means to handle the complexity resulting from the simultaneous occurrence of multitude and softness in design criteria. This is the problem tackled in this thesis.

The aim of the present thesis is to enable architects to arrive at designs that satisfy the criteria they deem relevant with certainty, allowing them to take their decisions with great awareness and confidence.

The premise of the present approach is that effective use of *computation* is an essential step to accomplish this aim. The significance of computation is because other means to handle the complexity introduce further uncertainty into the process. This is undesirable. For example, a certain 'way of thinking' or other mental strategies may have some effect [18]; however, clearly their effectiveness and validity is generally questionable. This is because they are inherently imprecise. Therefore we have no means to verify with some precision if a certain way of thinking is superior to another one, for example in the sense that one way is more likely to produce suitable designs than another one.

When we aim to use computation to deal with the design complexity, a bottleneck of particular importance is to handle the *soft* issues of design. These are considered to be incomputable conventionally. A significant step is to bring them into computation.

Having noted the gap in our scientific knowledge aiming for computational design enhancement, in particular considering the shortcomings of the existing computational design methods, the research question is: *How can soft criteria be brought into computation, and how can design solutions be found that satisfy the criteria when they are conflicting?*

1.4.2. Requirements dealing with the softness of criteria

Due to the complexity of architectural design a number of different aspects need to be taken into account when advanced design assistance is pursued; i.e., when computation is employed to arrive at desirable solutions as opposed to merely representing solutions.

We emphasize that architectural design involves soft requirements. If we wish to design a space that has a high *visual openness*, or a place where the columns and installations are *hardly perceived*, clearly such perception-based demands are not subject to being merely fully satisfied or not satisfied at all. This is because perceptual properties are attributes of a space that are not sharply defined. For instance let us consider a space with a large window overlooking a lake. In the space we are experiencing a certain level of visual openness. When, instead of the lake scenery, we were facing the wall of another building, clearly the openness we experience is reduced compared to the previous situation; however it does not vanish totally.

Next to their obviously soft nature, requirements that are based on visual perception are prime examples of soft design requirements. This is because architectural designs generally involve such requirements. After all our experience of architecture is essentially based on the visual awareness we have for the objects in our environment, such as the walls, ceiling, and furniture. Architecture addresses our esteem needs and needs for self-actualization essentially via visual perception properties of spaces. Therefore, perception is a matter of general relevance in architectural design. A common requirement is to ensure desirable views to the exterior, or to pronounce the significant stylistic elements of a building shape, so that these are more prominently noticed from the outside.

Architects make use of models to predict with precision the stability and strength of our buildings, how they behave under different wind loading, how much energy they will consume and what it will cost to build them. Up till now we have no means to predict with comparable precision visual perception aspects of spaces. This is because visual perception is a phenomenon that is vaguely understood and merely verbally described in the literature [19-22]. If we wish to let perceptual requirements play a significant role in design it is clearly necessary to treat perception in a more definitive form beyond verbal descriptions. This is the case in particular when opportunities need to be explored for weighing perceptual qualities against other criteria, such as costs, sustainability etc.

Due general relevance of perception aspects in architecture and the lack of available models to handle their soft nature, the first component of the computational approach pursued in this thesis is a model of visual perception. The purpose of the model is to let a designer assess the perceptual properties of his or her designs with great awareness and increased precision.

To illustrate the expected merits of a perception model let us consider the following example. Say we are designing a residential area, and we want that the houses of the area have living rooms and gardens arranged in such a way, that it is not easy to see into them from the other houses in the neighbourhood. In this case, when we contemplate a certain arrangement of houses as a solution we need to assess for every garden and living room, how strongly one visually perceives them from the other houses. This information should be provided by a perception model. In this way it is avoided that an architect has to spend time to assess the various spatial situations in the design ‘manually.’ This means he/she does not need to inspect renderings, or assuming multiple viewpoints in a VR environment. However, the computational perception assessments are also more precise, which is more important than the gain in efficiency. Namely they are made based on the same metric every time, ensuring that the imprecision of the result is minimal. In contrast to this, a human observer does not compare the perception from several viewpoints with such consistency. This is because the visual information of a scene is usually abundant. Therefore subtle perceptual differences among alternatives are likely to be overlooked, and it is problematic to identify a suitable direction to improve the design.

Increasing the precision of the information about a design is what also motivated the development of modelling stress and heat transfer in structural and thermodynamic analysis for example. In case of the present thesis the information is of perceptual nature. However, computational design assistance should not stop at provision of more precise information about the properties of a design. Such information should be made use of effectively, so that desirable designs are reached. To assist a designer in this task, an important step is to verify to which extent a design satisfies the design requirements.

This means that next to the perception model a second component is needed. This component should use the results from the model, and convert the perception information into degrees of satisfaction of the design requirements. In the example mentioned above we aim to design gardens that are hardly perceived from the houses around. Once we obtain the degree a garden is perceived we should determine to what extent this is satisfactory. The soft nature of such a requirement should be taken into account. This means we need to bring our personal preferences regarding how much perception we consider desirable into computational form. Also we should bear in mind how tolerant we are if the perception differs from what we wish to have ideally. Being able to take such imprecision into account is necessary because design requirements are often not totally fulfilled due to the conflicts with other requirements.

1.4.3. Requirements for dealing with the diversity of criteria

Clearly in design it is not enough to know the degree of satisfaction of individual requirements. Designs are usually performing well with respect to a number of aspects and poorly with respect to others. To decide which direction we should follow to reach most desirable solutions we need to have an idea about the overall 'goodness' of a design.

Therefore it is necessary to model the consequence of the simultaneous satisfaction of several requirements. Essentially this means establishing a model of the suitability of a design provided the requirements are satisfied in a certain way. The results of such a model should tell for example how suitable a design is when it is very functional, but moderately sustainable and has a low perceptual quality.

From this example it is clearly noted that in order to determine such an overall performance we need to know the relative importance among the criteria involved. However, in architectural design the relative importance among criteria is generally imprecisely known in advance [7]. This is because criteria such as *functionality* or *perception* aspects are soft in nature. Therefore a designer is hardly able to pin-point them exactly. Therefore a model for evaluating the performance of a design should be able to absorb the natural imprecision of such specifications. In other words such a model should be robust in the sense that the results it provides should not be severely influenced if a designer specifies the criteria with some imprecision.

A second requirement for performance evaluation is that an evaluation model should be able to deal with the partial satisfaction of requirements. This means it should be able to provide information about a design's performance although certain requirements are not fully met. This is necessary because design criteria are usually conflicting, so that they cannot be fully satisfied simultaneously.

1.4.4. Requirements for dealing with the multitude of solution options

Being able to evaluate the performance of alternative solutions, while taking the softness of associated criteria into account, is a significant target to accomplish. However reaching this goal is not enough considering the design context. After all the final purpose of design is to reach solutions that satisfy our criteria to the greatest extent we can achieve. This means we need to make use of the information we obtain from evaluating a design. Based on this information we aim to improve our design, so that it better satisfies our criteria. This means we need to consider alternative directions, modifying a design in such a way that its overall suitability is increased, and eventually becomes as high as possible.

In this respect it is significant to mention that the starting point of a design is important. It influences further directions a designer will be opting. As every design project is unique to some extent, so that design criteria are generally different from project to project, there is little information which solution to prefer. The information we have are boundary conditions for properties of the solution imposed by law, such as lot boundaries, maximum height, maximum energy use etc. From our experience we usually have an idea, what features a suitable solution should have. However, beyond this there is no information which combination of choices we make about the solution will yield solutions that satisfy our criteria to the greatest extent.

A design consists of many building components, and each one has a number of parameters to decide about. This means the amount of possible combinations that should be considered is enormous. Since we have limited time and little information on suitable solutions in advance, a computational method should explore many possible solutions generating designs that satisfy our criteria to maximal extent. This means we should employ a computer to investigate a vast amount of possible solutions in the feasible range, without a priori-preference for one solution over another one. In this way solutions matching the criteria are reached that were previously unknown. It is noted that design criteria are conflicting, so that the computations are supposed to arrive at the most suitable compromises. An architect is usually unable to specify in advance how important the criteria are relatively. This is because criteria involve highly abstract linguistic concepts, so that the consequences of committing about the importance among them are uncertain. Therefore a computational approach should aim to establish solutions, where such a commitment is not necessary. That is, an architect should be able to postpone this commitment until he/she is aware of implications this entails for the solutions.

1.4.5. Methodologies employed in this thesis

The requirements presented in the previous section are challenging to meet for conventional computational methods. This includes the methods from the domain of classical artificial intelligence. In the last decade a number of methodologies emerged in a sub-domain of AI known as *computational intelligence* (CI). They are also referred to as *soft computing* methodologies. These methodologies have features that allow meeting the requirements described above. In particular they are able to deal with partial satisfaction of requirements, imprecise requirements, such as perception-based requirements, and conflicts among them.

A number of applications of CI methods are reported in literature, and they proved to be successful handling complexity issues in a design process [23, 24]. Each of the CI methodologies has their distinct characteristics, while there is some overlap in approach and application [25]. Synergistic combinations of such methods are known as soft computing systems [26]. In such systems each method plays a distinct role, dealing with a certain aspect of the complexity of a problem. Such systems appear highly promising as a basis to deal with the requirements for a computational design approach put forward above. The main soft computing methodologies are neural systems, fuzzy logic, and evolutionary computation.

In this thesis the three methodologies are working together as a system. The system addresses the different sources of the complexity in design, providing a designer with solutions that match his or her criteria. In particular a model of visual perception is developed that evaluates different design alternatives with respect to certain perceptual properties. Based on the information from the perception analysis, the suitability of alternatives is compared using a performance evaluation model. By means of this model perception-based criteria are taken into account together with others. Finally, designs are generated that fulfil the criteria to maximal extent using the results from the performance evaluation.

1.4.6. Limitations of the approach

One of the required components for the computational design approach pursued in this thesis is a model of visual perception. It is emphasized that the large variety of possible environmental conditions encountered during design makes it desirable that the perception model is able to evaluate perception for a wide range of spatial situations. It is clear that a model that would tell about perception only for special types of rooms, or only for spaces with a special size or proportion, would not be suitable for architectural design purposes.

Also, preferably the model is not restricted to the perception of any person in particular or to any specific activity. One might develop a model that exclusively represents observers, who are searching for an object, so that they have a bias for a certain kind of object. Or a model might exclusively represent observers who have a certain kind of seeing habit. Such models clearly have a limited value for architectural design. This is because in general it is not known in advance what kind of pre-occupation a certain person has at a certain circumstance during his or her experience of a building.

Therefore the perception model pursued represents an observer, who is engaging in a perception activity that has some general validity. This case is an *unbiased* observer, who has no a-priori preference for objects in the environment. It is noted that modelling biased observers is treated in the literature [27]. In order to deal with a large amount of different spatial conditions the model concerns the general influence that geometry and colour have on the awareness of an observer for environmental objects. Other possible influences on the visual awareness, such as shape familiarity, task induced preferences for certain objects, or other contingencies, are excluded from the model. Establishing an unbiased observer model is comparable to assuming a certain representative loading condition in a structural analysis, although the particular condition may never occur in reality exactly. That is, it forms a reference.

A second limitation concerns the understanding of *design enhancement* used in this thesis. In the thesis design enhancement is understood in the sense that designs are obtained that meet the architect's criteria with increased certainty compared to the conventional case. Of course such provisions are independent from the question if the criteria are appropriately selected. In other words the architect is required to have the knowledge regarding appropriate criteria for the design at hand. The role of the computational system pursued is limited to that of a *partner* for the designer. In this partnership the designer defines criteria and solution boundaries, while computation aims to satisfy the criteria within these boundaries maximally.

1.5. Outline of the work

The computational approach developed in this thesis is based on a system consisting of three components. Each one deals with a specific source of complexity in design. The components are explained in separate chapters as follows.

Chapter 2 describes a model of visual perception that is used to assess perceptual properties of environments. This exemplifies a soft design aspect that is challenging to treat using conventional means. The results from assessing a number of soft requirements are then used to determine the overall suitability of alternative solutions. This is treated in Chapter 3, where a model for performance evaluation is presented. The results from this model are used in a design generation process, where they direct a computational generation process towards desirable solutions. This is described in chapter 4. The three components work together as a system making use of each other's outputs until desirable results are reached. The system is described at the end of chapter 4, and based on this the concept of a cognitive design approach is described. This concludes the theoretical part of the thesis. The cognitive approach is applied in two design applications, which are presented in Chapter 5. One of the design tasks was devised from a real scene for computation; the other one was an actual project. In the latter application the computational results are compared against the solutions obtained using conventional means. This is followed by conclusions. Appendix A gives details on the colour perception component of the perception model.

1.6. References

1. Sariyildiz, I.S., Ciftcioglu, Ö., Tunçer, B., Stouffs, R.: Knowledge model for cultural analogy in design and design education. In: Stellingwerff, M. (ed.): Local Values in a Networked Design World - Added Value of Computer Aided Architectural Design. Delft University Press, Delft (2003)
2. McDonough, W., Braungart, M.: Cradle to Cradle: Remaking the Way We Make Things. North Point Press, New York (2002)
3. Maslow, A.H.: A theory of human motivation. Psychological review 50 (1943) 370-396
4. Maslow, A.H.: Toward a Psychology of Being. John Wiley (1999)
5. Sariyildiz, I.S.: ICKT effects on the building culture and society. Joint Int. Conf. on Construction Culture, Innovation, and Management (CCIM), Dubai, UAE (2006)
6. Sariyildiz, I.S.: X-Dimensie in de Bouw: X-Dimension in Building - Inaugural lecture Publicatiebureau Bouwkunde, Technische Universiteit Delft, Delft (1994)
7. Lawson, B.: How Designers Think. Butterworth-Heinemann, Oxford (1997)
8. Hamel, R.: Over het Denken van de Architect. AHA books, Amsterdam (1990)
9. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological review 63 (1956) 81-97
10. Baddeley, A.D.: Working Memory. Oxford University Press, New York (1986)
11. Koile, K.: An intelligent assistant for conceptual design. In: Gero, J.S. (ed.): Design Computing and Cognition'04. Kluwer Academic Publishers, Massachusetts Institute of Technology, Boston, USA (2004) 3-22
12. Flemming, U., Woodbury, R.: Software environment to support early phases in building design (SEED): Overview. Journal of Architectural Engineering 1 (1995) 147-152
13. Eastman, C.M.: Automated space planning. Artificial Intelligence 4 (1973) 41-64
14. Gero, J.S.: Expert systems in computer-aided design. FIP WG 5.2 Working Conf. Elsevier Science Inc., Sydney, Australia (1987)
15. Alexander, C.: A Pattern Language. Oxford University Press, New York (1977)
16. Lighthill, J.: Artificial intelligence: A general survey. Artificial Intelligence: A Paper Symposium. Science Research Council (1973)
17. Russel, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, Upper Saddle River, NJ (2003)
18. Alexander, C.: The Timeless Way of Building. Oxford University Press, New York (1979)
19. Gibson, J.J.: The Ecological Approach to Visual Perception. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey (1986)
20. Foster, J.: The Nature of Perception. Oxford University Press (2000)
21. Palmer, S.E.: Vision Science. MIT Press, Cambridge (1999)
22. Levine, M.W., Sheffner, J.M.: Fundamentals of Sensation and Perception. Addison-Wesley, London (1981)
23. Caldas, L.G., Norford, L.K.: Genetic algorithms for optimization of building envelopes and the design and control of HVAC systems. ASME Journal of Solar Energy Engineering 125 (2003) 343-351
24. Shea, K., Sedgwick, A., Antonuntto, G.: Multicriteria optimization of paneled building envelopes using ant colony optimization. Intelligent Computing in Engineering and Architecture. Springer Verlag, Berlin Heidelberg (2006) 627-636
25. Hunt, K.J., Haas, R., Murray-Smith, R.: Extending the functional equivalence of radial basis function networks and fuzzy inference systems. IEEE Trans. Neural Networks 7 (1996)
26. Zadeh, L.A.: Fuzzy logic, neural networks and soft computing. Communications of the ACM 37 (1994) 77-84
27. Ciftcioglu, O., Bittermann, M.S.: From perceptual towards cognitive robotics in the framework of evolutionary computation. In: Pennacchio, S. (ed.): Emerging Technologies, Robotics and Control Systems. International SAR, Palermo, Italy (2009)

2. Visual perception modelling

This chapter presents a model of visual perception, where perception is explicitly defined and applied for perceptual analysis of spaces during design. The chapter is based on the publications [1-16].

Visual perception is an intricate phenomenon that is treated in various areas of science. An introduction to the topic is presented in section 2.1, where the significance of perception in architecture is highlighted. Existing works on perception are extensive. They are discussed in section 2.2 clarifying the scientific contribution of the present perception theory. The theory developed in this work is presented in section 2.3, where the concepts of visual *perception* and *attention* are put on a mathematical base. The theory explains common human vision phenomena, which are significant for architectural design. Section 2.4 describes implications of the probabilistic perception concept in the assessment of architectural environments. Next to geometry, colour plays a significant role in perception as well, influencing the degree of awareness an observer has for an object. This is treated in section 2.5, where the colour perception model is combined with the geometry based perception model. The resulting composite model is verified by means of two experiments.

Section 2.6 presents two implementations of the perception theory demonstrating the use of the theory for perceptual assessment of urban and interior spaces.

2.1 Introduction

2.1.1. Relevance of perception in architecture

Architectural design involves *hard* and *soft* aspects [17]. Examples for hard aspects are building regulations that have to be met. For example a regulation may restrict the building height to be maximally a certain height. Such aspects are referred to as 'hard' because they can be well defined, i.e. they do not involve imprecision or uncertainty in their specification. They refer to a classification of an object based on a quantity which is concrete, such as height, and based on an exact threshold defining the membership of the object in a class. As a consequence a hard requirement can be either met or not met, and there is no possibility for ambiguity on this matter. In contrast to this, soft aspects are ill-defined. That is, their description does involve imprecision or uncertainty. They may be partially fulfilled, and they refer to an association of an object with a linguistic concept, which is relatively abstract. For example a building may be said to have a 'high functionality,' which is a concept that is related with the physical properties of a building in intricate ways. Therefore soft aspects are generally difficult to treat. They have been considered incomputable conventionally. Prototypical examples of soft issues in design are aspects related to visual perception. Such aspects are ill-defined, subjective, circumstantial, and hence difficult to model. Goal of the present chapter is to demonstrate the treatment of visual perception, exemplifying the treatment of soft aspects in design.

Beyond serving as a prototypical example of a soft issue, perception is a relevant issue to tackle because it is important in architecture. Architecture addresses the basic human needs, such as needs for safety and shelter. Beyond that it addresses esteem-needs and needs for self-actualization of an individual [18]. This means a building reflects a person's personal values, and allows him/her to share them with others. Such high-level values are mainly addressed through perceptual qualities of spaces. An extrovert person prefers to live in a space that has an extroverted character. A person who is ordered needs to live in an orderly space, and so on. Beyond reflecting ones preferences, spaces may stimulate and cultivate qualities in oneself which one wishes to develop still. In practise, often users have little means or are given limited opportunity to express their preferences on perceptual properties of a building. When architects are aware of such desires and realize buildings that fulfil them, people benefit greatly from architecture. Such buildings are for example the expressive structures by Santiago Calatrava, the secluded residences by Tadao Ando, or the rich Baroque chapels by Balthasar Neumann.

Every person has their unique background and character. Therefore there is no general information regarding which perceptual quality is more or less suitable. An architect should determine this from case to case using his/her experience and knowledge. When an architect understands what spatial qualities may be suitable for a certain user, it becomes relevant how to realize them. For this it is necessary that the perceptual qualities of a space are well assessed. This way, alternative solutions can be compared and suitable solutions obtained.

The objects forming a space and their parameters influence the perceptual qualities. Knowledge on this influence is necessary to attain sophistication on the perceptual level. Our visual system is highly sensitive and even registers subtitle differences in our environment. Therefore it is desirable to evaluate perceptual properties of spaces with precision. This way, subtle differences can be taken into account, and we are able to verify how strongly a design solution satisfies our perceptual requirements.

2.1.2. Relevance of modelling perception in architecture

Architects and engineers use different computational models to analyze the properties of a design. For example computational methods are used for structural analysis, wind analysis, or cost analysis. The motivation behind the modelling effort is to obtain precise information in the evaluation of a design.

Although perception is an essential aspect of architecture and often a major factor determining the success of a design, there are no models to analyze the visual perception properties of built environment. This is because visual perception is a phenomenon that is not very well understood up till now. It is merely verbally described in the literature [19, 20]. Due to the *soft* status of the concept, it is problematic to consider perceptual properties together with other aspects, such as *cost* or *sustainability.* Perceptual requirements often conflict with other requirements. In order to reach suitable compromises it is necessary to consider the aspects on a common ground. This requires expressing perceptual properties in a more definitive form beyond verbal descriptions, which are vague. This way architects can determine the perceptual properties of their design with great awareness and precision.

For example in a hotel lobby we may wish to pronounce certain visual elements, such as the stairs and elevators. The aim may be for example to let persons entering the space notice these objects more strongly compared to other spatial elements. Other building components, such as the columns of the space, we may demand to be hardly noticeable, aiming for a generous spatial ambience. From common experience we know that it depends on many factors how strongly we notice an object. Clearly a prime issue is our viewpoint, the object's positions, sizes, materials etc. We also know that if we make small changes to an object's properties, we may strongly affect the perceptual properties of a scene and the effect the space has on its users. This can be seen in the residences by Corbusier, where the proportions and relative positions of the building elements are chosen with great care, so that modification of their relations may have negative influence on the spatial effect. However, next to the perceptual effect, the properties of design objects also influence functionality, safety and other aspects.

In design we wish to arrive at designs that satisfy at the same time perceptual, functional and other criteria. Therefore it is necessary to compare the fulfilment of our perceptual and functional requirements for a number of different design alternatives. This means we need a way to compare the perceptual difference among several scenes. Desirably the difference is obtained with high precision, in order to have much information regarding the potential improvement of the design. This comparison is a challenging task to accomplish for a human designer using conventional means, such as visual inspection of renderings of the situation or mock-ups. The results from visual inspection generally have a low precision. This is because it is problematic to retain the same judgement metric in mind when making the comparisons. In other words, when the perceptual differences are hardly noticeable, we are likely to miss them from time to time. This is because the visual information of an architectural scene is rich and visual processing occurs largely unconsciously [21]. It is difficult for a designer to remain sensitive and aware to subtitle differences in his perception of a complex scene. Therefore, for the same scene the same designer may make arrive at differing assessment results when he/she is inspecting the scene at different moments. In the same way judgements are expected to differ from designer to designer, because it is a rare contingency when their perceptual judgements coincide. The issue becomes problematic in a cooperative design context, where different designers work on the same design together and should communicate about the perceptual properties and associated perceptual requirements.

In order to deal with perceptual properties of spaces with increased precision, it is necessary to establish a model of visual perception. This means perception should be defined in a form that is amenable for computation.

2.1.3. Difficulties to model perception

Modelling visual perception is challenging mainly because it involves not only the eye, but also the brain. The final “seeing” event occurs due to brain processes.

The complexity of the biological vision system is responsible for a common phenomenon we encounter practically every moment, while it may remain unnoticed: We *overlook* items in our environment, although they are visible to us. Overlooking refers to the inability to remember visual information, although the information was processed by the visual system. This common phenomenon is particularly noticeable in environments with many objects. When we view a scene for some time, and then try to remember the items of which the scene consists, we may not remember all of them. Although light reflected on the objects located within our visual scope and was processed by our visual system, we may not be sufficiently aware of all the objects to remember all of them. We refer to this phenomenon as *uncertainty* of human vision.

The uncertainty of human vision is a characteristic property of human vision. It occurs presumably due to the way the human memory is built up and facilitates consciousness dealing with ample environmental information. In this respect the human vision system clearly differs from an optical system like a camera system. The image of an environment acquired via a camera is permanently stored in the memory system of the camera. In case of human vision, memory is much more elusive. When we view a scene we are not aware of the items in our environment to the same extent. Our degree of awareness may be too low for some items, so that we do not consciously realize their presence and are unable to remember them. Still we did have some degree of awareness for such an object, and as the awareness was apparently too low for us to be conscious of the object, we can suspect that we *unconsciously* perceived the object. An overview of the evidence for unconscious perception can be found in the literature [21].

In scenes consisting of only few items, the degree of awareness we have for objects is generally high, i.e. we are consciously perceiving the objects. Still it should be noted that our degree of awareness generally differs from object to object. This may be difficult to verify experimentally. For example in a room that contains merely one item, e.g. a desk, it is unlikely that the desk will be overlooked. However, this may occasionally happen. Considering the same experiment we can expect however that one always remembers the room itself. In this case we were more aware of the room than the chair.

The commonly experienced uncertainty of “seeing” or *graded object-awareness* has never been explained precisely. This is mainly due to the complexity of the brain processes involved in human vision. These processes are not sufficiently known, yet. Clearly we may be more aware of an object in our environment, when it is more significant to us. This may be the case for example when we need the object to accomplish a task for example, or when it is unexpectedly present and may form a potential danger. However, it should be clearly noted that regardless of such a bias the geometric properties of the environment influence perception. We can easily verify this considering that we are usually aware of an object when it is located nearby, and in front of our eyes; however we may easily “overlook” the same object if it is placed at great distance. This is valid even when we are looking for this object. We all know the influence of geometry intuitively. For example when we leave a note for a person we place the note in such a way that when the person will enter the room he/she will have greatest chance of seeing it. We will do this intuitively right to some extent, however it is difficult to explain precisely what we do, and what the difference would be if we did it slightly differently.

The influence of geometry plays a particularly significant role during the phase known as *early vision* in the literature [22-25]. This is the process, in which an observer builds up an initial comprehension of his/her surroundings. The initial awareness of the environment is decisive for later stages in vision and ensuing cognition. In early vision, awareness is unbiased. This means that the observer has no a-priori preference for an object in the environment over another one. However, the influence geometry plays in this process is not precisely known up till now.

Goal of the present study is to gain insight into the role geometry plays in unbiased perception. That is, it is aimed to model how it influences our awareness of objects. This is done, so that perception related design requirements can be put forward with increased precision, compared to the conventional case.

In this chapter a novel model of early human vision is described. The model quantifies the degree an observer becomes aware of environmental objects. In this way the commonly experienced *overlooking* of visible information in perception, or *graded awareness* of objects, is explained. The model is based on probability theoretic considerations, so that it can duly deal with the complexity of the human vision process. To the authors best knowledge this is a unique probabilistic approach to perception. It is noted that the model is not *statistical* as this is conventionally the case in literature with Bayesian inference [26].

Due to the diversity of existing approaches related to vision and visual perception, before developing the model a reasonably comprehensive introduction is provided. This is to make explicit the contribution of the present research.

2.2 Existing approaches

2.2.1. Philosophical considerations

Perception is a deeply involved concept. This is indicated by its relation with a fundamental philosophical question. This question is whether the reality is independent of the human mind [19, 27]. In the philosophical discourse concerning the issue three major positions emerged. First is *idealism*. Philosophers taking this position put forward the concept of *mind-dependence* of reality. They consider reality as an entirely mental phenomenon. Second is *direct realism.* From this position the existence of the physical world is considered as logically independent from the human mind. Third is *representative realism.* Perception of a physical item is considered to be mediated by a mental item, which represents the physical one. Verification of the correctness of these positions is difficult due to the complexity of perception. Due to the complexity of the physical phenomenon, perception remained subject to verbal description. Some examples of such descriptions are "Perception refers to the way in which we interpret the information gathered and processed by the senses," [28] and "Visual perception is the process of acquiring knowledge about environmental objects and events by extracting information from the light they emit or reflect" [29]. Verbal definitions are helpful to understand what perception is about. However, they do not hint what it is exactly, and how to treat perception quantitatively. Demystification or precise description of perception is significant, so that the application of the concept in architectural design is enhanced.

2.2.2. The image processing approach

In vision science visual perception has been considered as the reconstruction of a three-dimensional scene from two-dimensional image information [29-32]. This approach is referred to as *image processing* approach. It attempts to mimic the biological and neurological processes involved in vision, with the retinal image acquisition as starting event. This is done by taking a two-dimensional image as a starting point in the model.

The approach is based on modelling the sequence of eye/brain processes that treat the retinal image by means of algorithmic counterparts. This is a formidably challenging endeavour. The reason is that these processes are complex and conditional, since brain processes play a decisive role. The involved brain process components as well as their interactions should be known with certainty when a deterministic approach is followed to model perception. Despite significant progress on this matter [33] the existing knowledge is currently not sufficient. Therefore deterministic approaches do not explain the characteristic uncertainty of perception, which is commonly observed. Well-known observations of visual phenomena, such as *depth from stereo disparity* [34], *Gelb effect* [35], *Mach bands* [36], *gestalt principles* [37], *depth from defocus* [38, 39], *depth from focus* [40] etc. reveal components of the vision process. However, it is unclear how they interact in human vision. In particular it remains mysterious, how the combination of the modelled components results in the uncertainty of perception.

2.2.3. The ecological approach

In his *ecological* or *direct* approach to perception Gibson [20] identifies the shortcomings of the image as a starting point to model perception. He proposes to consider perception as *direct pickup* of environmental information. The pickup is by means of successive eye fixations. The concept is appealing. However, neither Gibson, nor his followers [41] provide precise insight how to model the eye fixations and the pickup of the information. Therefore perception remains a verbally defined phenomenon. Ecological optics explains *why* certain properties of the visual system are necessary. However it does not explain precisely *how* the visual system works. This is in the sense that it remains mysterious how the visual system yields awareness of an environment, and what role the geometry of the environment plays in this process. In particular the commonly experienced overlooking of visible information is not explained. However, the author of the present work agrees with Gibson's view, which reads "in contrast to its use in communication theory information should be construed as specificity of the useful rather than as the uncertainty of the specific. This sense of information is the best way to understand how perception could control behaviour" [41]. In the present work information obtained through perception is construed as specificity, which is the *awareness* of the environment. The imprecision involved in the cognition of this awareness is subjected to probabilistic modelling, as this will be described in the next section.

2.2.4. Pragmatic approaches

Inspired by Gibson's work, in the context of event perception research a concept related to human vision has been introduced. This concept is known as *isovist* [42]. An isovist is the array of the distances that span between an observer's viewpoint and the surfaces of the environment that are visible from the viewpoint. Isovists are used for spatial analysis in urban, architectural and landscape design, as well as in psychology research [43, 44]. An isovist can be considered a model of the final paths light takes reaching the retina of an observer. It is not a model of visual perception. Perception involves the processing of visual stimuli yielding the mental realization of environmental objects and events with its characteristic uncertainty. Works based on isovist omit this essential aspect of perception.

In the domain of geographical information science a method termed *visibility graph* has been introduced [45]. This is a graph containing as nodes a number of locations that are mutually visible in an environment. Visibility graphs are used in urban and architectural design for spatial analysis [46]. They are used to approximate the volume defined by an isovist. A visibility graph can be seen as a model of the space spanning between an observation point and the visible surface forming the environment. Clearly, it is not a model of visual perception. This is because it does not model the vision process with its characteristic uncertainty.

In the domain of virtual reality perception models are developed. In [47] the *visual acuity* of the human eye as well as the *minimum eye resolution* are modelled. Positions in the environment, which are located within a limited central region of the visual scope, are considered to be "perceived", while objects located outside this region are "not perceived." The approach ignores the common phenomenon that we overlook items in our visual scope, even when they are located within a limited central aperture in our visual scope, although such overlooking may be unlikely to occur.

2.2.5. Brain research and psychology

Brain researchers trace visual signals as they are processed in the brain. A number of achievements are reported in the literature [48-52]. However, due to complexity there is no consensus about the exact role of brain regions, sub-regions and individual nerve-cells in vision. Therefore it is unclear how they should be modelled. The neuroscience models are all different. This is because there is enough modality of the brain to accommodate all conclusions directed by the research. Therefore the approaches are not unified, and a particular brain process like *perception* is not explained on a common ground.

The difficulty to model perception is increased by the involvement of attention mechanisms in perception [53]. Experiments showed that what is remembered from a scene depends on attention [54, 55]. However, despite experimental research on attention and perception it remains uncertain what the concepts exactly are, and how they can be modelled quantitatively. The uncertainty originates from the difficulty to distinguish precisely among the various factors involved in a viewing experience. The factors range from the geometry of the environment, task induced bias, to personal preferences. Therefore identification of the influence attention has on the visual processing is formidable. This is due to the complexity of the brain processes. When a visual stimulus is processed in the brain, apparently several parts of the brain are responsible to select, prepare, and maintain certain neural activity. This interaction among the parts is responsible for the fact that an observer is more aware of certain information [53]. These parts are connected with many other regions throughout the brain. Due to the complexity it remains uncertain to what extent a neural activity in the visual cortex stems from the original stimulus, and to what extent it is due to attention related processing.

In the domain of psychology two essential views emerged as to the sequence of information processing occurring in perception. One view is termed *associationist* view. It states that perception of objects occurs through combining more elementary sensations. That is, first elementary sensations occur, and then they are integrated forming a whole, meaning the integration is understood to be an object. In contrast to this Gestalt theory claims that first one becomes aware of the whole, and only afterwards, if necessary, an observer analyzes the whole. In this respect the perception theory that is going to be explained in this chapter takes the associationist view in the sense that when an object is perceived it necessitates that attention is paid to the parts constituting the object. However, in the present perception theory, in contrast to existing work taking the associationist view, it is well defined what attention is namely through a mathematical model bearing minimal ambiguity. In contrast to this, verbal statements, such as “attention is a tool to adapt what we see to our current needs,” [56] give no clue on what attention is exactly, and therefore also no precise instruction on how to measure it.

A concept that received significant attention in the domain of psychology is the concept of *feature integration* [57]. Feature integration is based on experiments concerning higher level tasks, such as visual search and texture segregation. In these tasks objects may appear at different locations on a display. The results indicate that for increasing display size the time required in average to perform perceptual tasks increases as well. Also, from experiments it is found that attention involves a *scope* [56], and when this scope is narrow, object features can be obtained and correctly combined with greater certainty than in case the attention is distributed over a larger amount of objects. However, the role the size of display and scope play on the perception is not precisely explained in these works. This is due to formidable complexity of the vision process. Both phenomena of feature integration may be substantiated based on the probabilistic perception theory that is described in this chapter. Namely concerning the effect of display size it is noted that according to the probabilistic perception theory, increasing display size reduces the attention, i.e. the perception per unit length of display size decreases. This will be clear from Eqs. (2.1) and (2.10), as in the latter equation the reciprocal value of the visual scope’s appears as a factor, thereby scaling attention. Compensation for reduced attention is what may explain why more time is required to obtain sufficient perception of the features for accomplishing the task. Concerning incorrect combinations of features, so called conjunction errors, are also observed. These may also be explained due to low degree of perception of an object. Then the objects' features and properties are obtained with a lower probability, i.e. the observer is less aware of them, possibly resulting in conjunction errors.

Generally it can be stated that due to the high complexity in the phenomena experimental psychologists are investigating, it is desirable to reduce the complexity. That is, we need a basic understanding with consensus to comprehend the phenomenon better. For example the reason for a certain relationship between display size, and response time in the experiments can be understood with greater precision than what is known to date, since the influence of geometry may become already theoretically known, namely using a the perception theory that will be explained shortly. This way the experimental findings may be substantiated and the precise influence of feature similarity on perception regarding various feature dimensions may be distinguished with greater precision. Making use of the probabilistic model the complexity of experimental analysis is reduced, so that identification of the influence of detail features on perception, such as shape similarity, becomes more precise. We may also expect that conjunction errors may be more accurately predicted, and also the classification of features as 'separable' and 'integral' feature becomes sharper.

It is emphasized that without knowing the effect of unbiased early vision it is formidably challenging to discern the specific effect of detailed object features on perception, such as task specific or ecological priming for instance.

It is further emphasized that the influence of geometry is not identified in the works on attention mentioned above. Without precise understanding of this influence, distinction among this effect from other factors, such as task specific bias or personal preference, is problematic. This means, without a model of the *unbiased* perception its influence in a biased perception event is uncertain. This entails that the final seeing event is not uniquely or precisely modelled, and the attention concept remains ill-defined. Since attention is ill-defined, ensuing perception is naturally also merely ill-defined. The relation between environmental stimulus and the mental event of perception remains unknown.

As a recapitulation of the previous section we note that visual perception and related concepts have not been exactly defined but qualitatively considered among the general vision related issues. After all it appears that the more details are discovered about the functionality of the brain, the more challenging it becomes to grasp the relation between a stimulus and a corresponding mental realization of it. Therefore, the perception phenomenon is not explained precisely or quantified. Consequently the use of the perception concept in architectural design is limited to application of personal experience or knowledge obtained from experimental data.

2.3 A novel theory of visual perception

2.3.1. Objective

Vision is our essential source of information when we experience our environment. Based on vision, many related concepts that are derivatives of vision can be defined. Such concepts are visual perception, visual attention, visual openness, and so on. However, these concepts are only verbally described up till now. Therefore there is no consensus in literature as to the definitions of these concepts, although they roughly coincide in essence.

Architects need to treat perception with high precision in order to achieve sophisticated designs. This means a model of perceptual properties of environments should be established. Such a model can be used as a tool by architects who pursue certain perceptual qualities in their designs. Due to the large amount of possible spatial scenarios encountered in a design it is necessary that a model is established that describes the general perception influences. It should not be restricted to specific environments. This means a model should be formed that describes the fundamental influence of the common design parameters, such as position, size, and colour on perception. Such a model currently does not exist. Therefore visual perception related criteria in design are subject to imprecise assessment at large. Let us consider we wish to evaluate a design in terms of its perceptual properties. For example we would like to know to what extent certain design objects are perceived from a specific viewpoint in the space. Such visual assessment bears imprecision, when we use visual inspection as the base of it. The imprecision is particularly in the sense that our assessment may vary significantly from case to case and from designer to designer. This is because a large amount of factors influence our awareness. In particular our attention is influenced by circumstances, previous considerations etc. Also we are easily overwhelmed by visual detail when we inspect a visually rich scene.

Next to the imprecision, another drawback of conventional perception assessment is that it may be rather time consuming. This is because the appropriateness of a design is usually assessed from multiple viewpoints and on several perceptual criteria. This issue becomes more severe when a large number of design alternatives are subject to comparison.

Having noticed that, this gap in our knowledge the present work endeavours to establish a model of unbiased vision. The aim is to model the dependence of visual awareness on fundamental environmental properties. In particular the properties concern the geometry of the environment in the first place. As result from establishing such a perception model it is foreseen that several elusive concepts like visual perception and visual attention are no longer elusive but subject to quantification and computation. Based on such a model, perceptual requirements can be put forward in precise form. This way they are subject to computational treatment. This implies that perceptual aspects can be weighed against conflicting requirements with great awareness.

2.3.2. Approach

Human vision involves image acquisition with the eye and interpretations in different regions in the brain. This process is formidably complex and details of it are not known. Therefore, when deterministic methods are used modelling the process is formidably difficult. Therefore, in the present section a novel approach to model perception is presented. In this model visual perception is put on a mathematical foundation. This is accomplished by means of the probability theory. The probability theory is particularly suitable to handle the high complexity and uncertainty involved in the human vision process. The main principle of modelling by probability theoretic considerations is to absorb complexity and uncertainty into probability. With respect to human vision, a probabilistic model is appealing, since it can absorb the complexity of the biological information processing by the probability theoretic considerations. The result is a match between model outcome and the perception phenomena we commonly experience.

In the present approach the components involved in human vision are modelled as a whole system instead of modelling each component individually. Thereby the model bridges from the environmental stimulus to its mental realization. The model we are about to consider has ample space to accommodate neuroscience, biology or other interdisciplinary aspects of perception without any restriction. At this point the definition of the concepts involved in vision is not elaborated further. This will be established naturally as result of the vision model, which is described in the next section.

The probabilistic modelling approach presented here can be seen as a unifying model. This is because it unifies synergistic visual processes of human. The processes include physiological and neurological processes, as well as philosophical aspects of vision. Interestingly this is achieved without recourse to neuroscience or biology.

The work concentrates on the influence of geometry on the human vision process, where an observer builds up an *unbiased* understanding of the environment. This means an observer has no a-priori preference for an object in the environment. Such a bias may be due to a task the person is about to accomplish, or a general personal preference. In many instances we expect some degree of bias in visual perception. However, it should be clearly noted that it is significant to model the unbiased case. When the unbiased case is known there is a basis to model biased cases with greater precision, compared to the situation where the unbiased case is unknown.

Explicitly the model gives two advantages:

- The perception and related phenomena in early vision are understood in greater detail, and some common reflections about them are substantiated.
- The model can be effectively introduced into architectural design, since perception is quantified a probability.

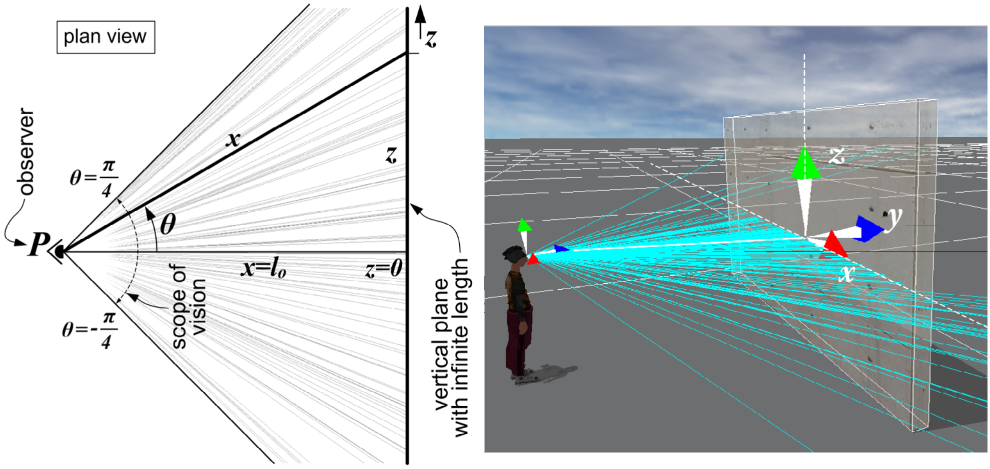
This is accomplished by not dealing explicitly with the complexities of brain processes or neuroscience theories, about which more is unknown than known. The vision model is derived based on common human vision experience explaining the causal relationship between vision and perception in unbiased human vision. For this reason the presented vision model modestly complements existing works in the sense that they can eventually be coupled to the output of the present model. The novel model introduced here is based on a premise, which is presented in the following section. It explains a number of perception related phenomena that we commonly experience as result of our vision process. In this way the integrity of the results from the model are conformed through the practical implications. It is emphasized that the approach is based on a premise, and the implications of this premise are studied. This approach is oppositional to conventional approaches in psychology or brain research, where a model is developed essentially from collected data.

From the introduction above, it should be emphasized that the research presented here is about to demystify the concepts of perception and attention in human vision from their verbal description to a scientific formulation. Due to the complexity of the issue, so far such formulation is not reported.

2.3.3. Probabilistic theory of perception

We start modelling the perception process using a simple, yet a fundamental geometry. This is shown in figure 2.1. For the sake of simplicity we first consider perception in a 2-dimensional space. That is, the vision of an observer is considered in the x-y plane in figure 2.1. In figure 2.1b an observer is facing and viewing a vertical plane from the point denoted by *P*. Through vision, the observer is able to receive visual information from all directions within his/her vision scope.

In the present case we model *unbiased* human vision. This means the observer has no preference for any direction in the scope. In other words he/she pays attention equally in



(a) (b)

Fig. 2.1 Plan view of the basic geometric situation of perception (a); perspective view of the same situation (b)

all directions. In mathematical terms this is modelled by associating an equal probability with any differential angle portion in the visual scope. That is, the observer pays attention to all locations on the plane with uniform probability density concerning the angle θ defining the viewing direction. This premise is sketched in figure 2.2. The premise ensures that there is no visual bias in perception. The unbiasedness concerns the direction, from which environmental information is obtained. Therefore the model is a model of unbiased vision, i.e. it is a model of early vision, namely before a bias is exercised. This is in contrast to biased vision occurring in visual search for instance. The latter phenomenon occurs after unbiased vision has occurred. One notes that unbiased vision is an idealized case. In many instances vision may be biased. A model for biased vision is presented in another work [58].

Individuals may have certain a-priori preferences for certain visual information, when they are performing a task, where they deem certain environmental information of particular significance. It is noted that in the following computations are presented for the unbiased case, however probabilistic perception models can also be established for biased vision based on a general probabilistic perception theory [58]. Therefore the computations presented form a standard observer model. This model has merits from the viewpoint that it is generic suiting general perception measurement purposes in design.

Unbiased vision occurs during the phase known as *early vision*. Before an observer has information about the environment he/she is about to experience visually, there is no base to justify any preferred direction in the view, in which attention should be paid. If we were able to specify in advance a certain direction in which we expect to have more environmental information, the bias must be based on previous information about the

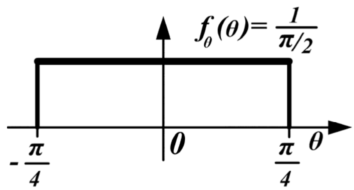


Fig. 2.2 Premise of the probabilistic perception model

scene. That is, modelling an unbiased observer is equivalent to the case that the environment is unknown to the observer. This means there is no information based on which the observer may exercise a bias. Hence visual attention is paid equally in any direction within the scope. This is the premise of the model, and it is based on basic human vision experience. It is not an ad hoc approach trying to explain a certain vision phenomenon. Instead this is a starting point from scratch to arrive at some results through further derivations, which follow. Therefore the justification will follow afterwards.

The premise shown in figure 2.1 entails that the probability to obtain environmental information is the same for each single differential visual resolution angle $d\theta$. This means that the probability density function (pdf) belonging to the angle θ is uniformly distributed. This pdf is given by Eq. (2.1). Assuming the scope of vision is defined by the angle $-\pi/4 \le \theta \le \pi/4$, the pdf sketched in figure 2.2 f_θ is given by

$$f_\theta = \frac{1}{\pi/2} \tag{2.1}$$

To illustrate the implication of the premise a number of "sight-lines" are shown in figure 2.1. These are generated based on the uniform probability density given by Eq. (2.1). Each of the lines represents an event of paying attention to a certain location on the plane.

The angle θ is a *random variable* in the terminology of the probability theory. Since θ is trigonometrically related with each point on the plane, the distance x, which is indicated in figure 2.1, is also a random variable. The pdf $f_x(x)$ of the random variable x is computed as follows.

$$tg(\theta) = \frac{x}{l_o} \tag{2.2}$$

In Eq. (2.2) l_o is the distance between observer and object.

Now we aim to identify the implications of the premise in Eq. (2.1) with respect to the object viewed. That is we are looking for the pdf along x direction in figure 2.1. For this purpose the theorem on the *function of random variable* [59] is applied: To find $f_x(x)$ for a given x we solve the equation

$$x = g(\theta) = l_o \, tg(\theta) \tag{2.3}$$

for θ in terms of x. If θ_1 , θ_2 ,...., θ_n , .. are all its *real* roots,

$$x=g(\theta_1) = g(\theta_2) = \ldots\ldots = g(\theta_n) = \ldots.$$

Then

$$f_x(x) = \frac{f_\theta(\theta_1)}{|g'(\theta_1)|} + \ldots + \frac{f_\theta(\theta_2)}{|g'(\theta_2)|} + \ldots + \frac{f_\theta(\theta_n)}{|g'(\theta_n)|} + .. \tag{2.4}$$

Clearly, the numbers θ_1 , θ_2 ,...., θ_n , .. depend on x. If, for a certain x, the equation $x= g(\theta)$ has no real roots, then $f_x(x)=0$.

According to the theorem above we write

$$g'(\theta) = \frac{dx}{d\theta} = \frac{l_o}{\cos^2(\theta)} = l_o(1+tg^2(\theta) = l_o(1+\frac{x^2}{l_o^2}) \tag{2.5}$$

$$\theta_1 = arctg(x/l_o) \tag{2.6}$$

$$f_x(x) = \frac{f_\theta(\theta_1)}{|g'(\theta_1)|} = \frac{l_o}{\pi(l_o^2 + x^2)} \quad (2.7)$$

for the interval $-\infty \leq x \leq \infty$ [8]. For this interval, the integration below becomes

$$\int_{-\infty}^{+\infty} f_x(x)\,dx = \frac{l_o}{\pi} \int_{-\infty}^{+\infty} \frac{dx}{x^2 + l_o^2} = 1, \quad (2.8)$$

as it should be as a probability density function. For the interval

$$-l_o \leq x \leq l_o \quad (2.9)$$

$$f_x(x) = \frac{2}{\pi} \frac{l_o}{(l_o^2 + x^2)}, \quad (2.10)$$

so that

$$\int_{-l_0}^{+l_0} f_x(x)\,dx = \frac{2\,l_o}{\pi} \int_{-l_0}^{+l_o} \frac{dx}{x^2 + l_o^2} = 1, \quad (2.11)$$

as one has to expect. The plot of $f_x(x)$ as a function of x is shown in figure 2.3. From Eqs. (2.7) and (2.10) it is clearly seen that the shape of the resulting pdf depends on the distance l_o between observer and object. This dependence is sketched in figure 2.4a and 2.4b.

Comparing these figures it is seen that for short distances from the object the pdf becomes narrower and more amplified in the forward direction.The pdf becomes wider and more 'flat' for an object at a greater distance. This is seen from the density of vision rays arriving at the object per unit length in *x* direction. Clearly this density is lower in the distant case compared to the near case. The meaning of the lower density of rays arriving at the object is that an unbiased observer will be less aware of such areas. If the density is high at a certain part of the object he will be more aware of that part. Being more aware entails that the observer will be able to remember the object clearer. That is, he/she will remember more details of the object. We note that the pdf is a Cauchy function, which resembles a Gaussian function, and the width of a corresponding Gaussian is approximately equal to l_o [14].

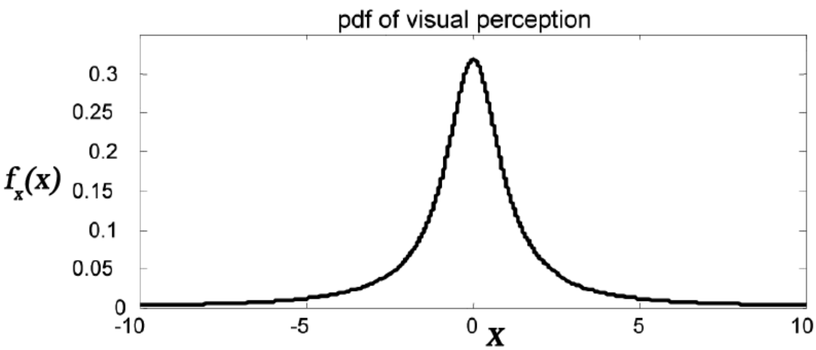
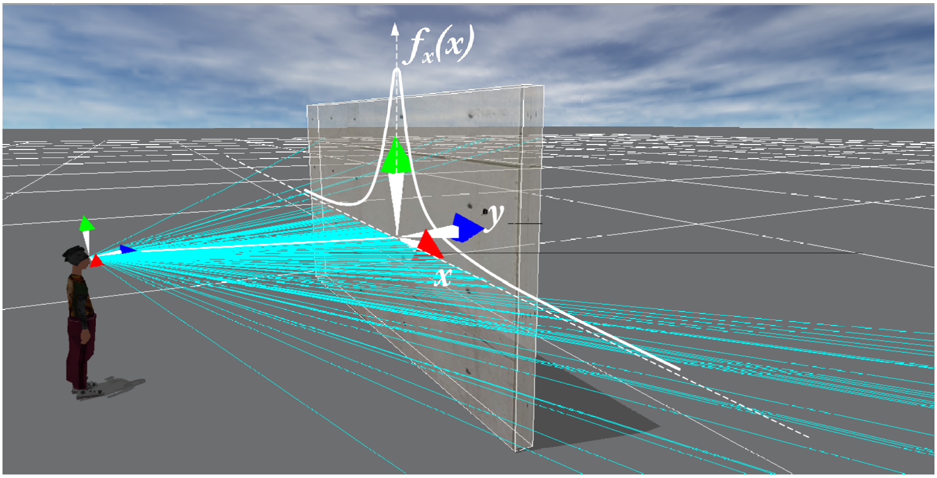
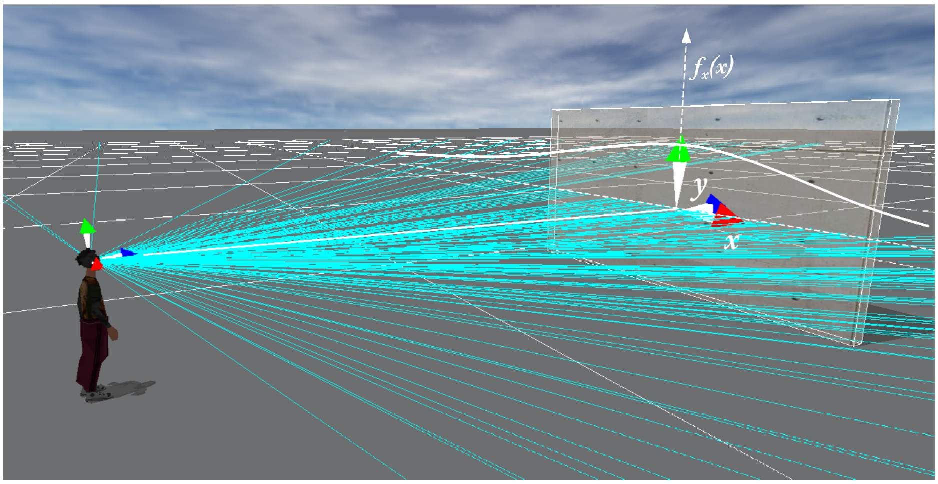


Fig. 2.3 Probability density function of unbiased perception for $l_o=2$.



(a)



(b)

Fig. 2.4 Unbiased attention for an object at a near distance (a); at a far distance (b)

2.3.4. Terminological definitions

Based on the results of the basic theoretical considerations above, some vision related concepts can be terminologically defined. That is, they can be expressed using mathematical terms, as follows.

V*ision* is the ability to see, i.e. it enables becoming aware of environmental objects. All vision-related concepts, such as attention, perception, visual openness and others are the sub-areas of vision. This means, vision is the essential necessity for the further considerations in these sub-areas.

The probability density function $f_x(x)$ derived in the preceding subsection corresponds to *visual attention*. This is a well-known concept in cognitive psychology [56, 57, 60, 61]. However, up till now it was not formally defined with precision. For example a definition is given by [62]: "Attention is … mental energy, or resource, or capacity enabling processing," and the author admits that "unfortunately there is little if any agreement about how … the mysterious, limited resource or capacity is to be defined." Another example of a definition is given by [56]:”Attention is a tool to adapt what we see to our current needs,” However stating the purpose of a phenomenon does not specify what it is exactly. It is interesting to note that another example of a verbal definitions of attention is given by [63]: “Everyone knows what attention is. It is the taking possession of the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization and concentration of consciousness are of its essence.” This description is valid; however, it is unclear how to model the focalization and concentration yielding the act of “taking possession of the mind.” Comparing the latter description of attention with the former two it is noteworthy to point out that within one hundred years of research on attention few progress has been made with respect to the precision of our understanding what attention is.

From the derivations above it is noted that visual attention is given by a probability density within an infinitesimally small spatial region, and per unit length interval. Integration of the probability density function $f_x(x)$ within a finite length interval yields a probability. This probability quantifies perception. The result is eventually referred to as *unbiased vision probability*. The probability is associated to the event that an observer is conscious of an object occupying the respective length interval in the environment for which the attention is paid, i.e. integrated. This means perception is modeled as a probabilistic event. It is noted that the definition of attention and perception given above may not exactly coincide with those used in the psychological literature. This is due to generally vague and verbal definitions in the latter case. However, the mathematical formulations presented above are decisive. They are subject to adoption into diverse fields. Among these fields most prominently are architectural design and robotics [5].

Integration of visual attention in a finite small interval, where $f_x(x)$ is approximately constant, gives unbiased perception as probability *P*

$$P = \int_{\Delta x} f_x(x)\, dx \cong f_x(x)\,\Delta x \,, \tag{2.12}$$

so that the probability becomes proportional to attention in this case. For the forward direction of viewing the infinite plane where $\theta \approx 0$ the vision probability becomes

$$P \approx \frac{2}{\pi}\Delta\theta \,, \tag{2.13}$$

where $\Delta\theta$ is the vision resolution of the human eye. Although in the theoretical development the region over which attention is integrated to yield the early vision probability can be infinitesimally small, in the practical case the region is restricted by the human eye resolution, which is defined by the differential angle $\Delta\theta$.

In an environment with abundant objects we can interpret the probability quantifying the perception literally as the *probability to see*, where seeing means an object can be remembered by the observer. That is the object was not *overlooked*. Essentially overlooking refers to the event that an observer was not *aware enough* of the object for being able to remember it. It is remarked that the inability to remember is despite the fact that an object’s visual information was processed by the observer’s visual system.

Understanding the probability as expression of an observer’s *degree of awareness* for an object is a more general interpretation of the resulting probability than to interpret it as ‘probability to see’. The former interpretation becomes apparent when we consider an environment consisting of merely a few objects. This way remembering the objects appears to be certain. In this case the interpretation of the resulting probability as *probability to see* is still valid, however it becomes difficult to verify. This is because the verification may

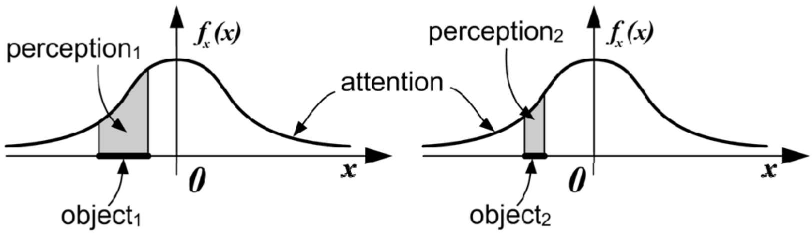


Fig. 2.5 Sketch showing the different degrees of perception for two objects with different size positioned approximately at the same location along the *x* axis given in figure 2.1

require an unrealistically high number of vision experiments. Also the experimental conditions should be the same every time, which is difficult to realize. In case there are few environmental objects, it is clear that the *degree of awareness* we have for the objects in our view is usually not the same for different objects. This is despite our ability to remember all of them.

2.3.5. Corroboration with common experience

The results from the above computations are in conformity with the common human experience in early visual perception. Namely, for the plane geometry shown in figure 2.1 the visual attention and thereby vision probability is strongest along the axis of the cone of vision, relative to the side directions. This means that an observer is more aware of details of the plane in frontal direction, i.e. these details have a greater chance to be seen, and hence remembered, than the side parts. One can easily verify intuitively the results from the probabilistic vision model above using basic vision experience.

For instance, in a scene, when one considers only the geometry and not any other aspects like illumination or colour, comparatively an object with bigger dimensions is more likely to be seen. Therefore early vision probability of this object is high. If there is more than one object in the scene, then the final seeing action for these objects will be merely subject to probability. This means it will be subject to perception. This is because the attention during early vision, i.e. the discrete probability density, would be distributed among these objects. The bigger object would be more likely to be seen in this case. This is shown in figure 2.5.

In case we position an object at different places along the *x* axis, the closer the object is to *x=0*, the greater is the vision probability of it is. This is shown in figure 2.6.

As the vision probability is obtained through integration of attention over a particular spatial domain, the perception model is closely related to gestalt principles. Gestalt theory addresses the phenomenon that an observer experiences a number of visual items as belonging together, so that they form another visual entity, i.e. a group. The group of is referred to as *gestalt*. With respect to the role of gestalt theory for visual perception it should be noted that an observer executes *grouping* based on perception, as introduced here. This entails the observer should have mentally realized the elemental visual items to be grouped beforehand. The number of domains and their shapes in a scene to be integrated is subject to probabilistic considerations. In these considerations the elementary probability of a gestalt quality denoted by *p* can be assessed by means of the perception considerations presented here. Further computations can be carried out with underlying probability theory [37]. In this way seamless coupling of the gestalt theory to the perception model presented here is an interesting relevance.

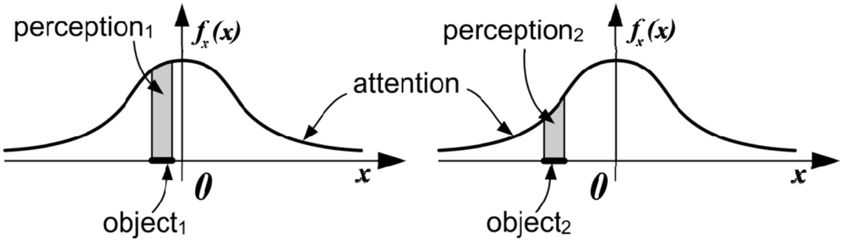


Fig. 2.6 Sketch showing the difference in unbiased perception of two objects with the same size positioned at different locations along the *x* axis in figure 2.1.

2.3.6. Measuring perception by exponential averaging

For a complex environmental geometry, to obtain the pdf specifying the visual attention in analytical form becomes infeasible. In particular the complexity refers to a scene that consists of a large amount of surface patches that are oriented in different directions with multiple occlusions occurring among them. To handle the geometric complexity an approach is used, where the degree of perception is approximated by means of probabilistic ray-tracing. The results from the ray-tracing are integrated in real-time. In this approach a virtual human is used, where elemental perceptions are obtained using rays, which are termed *perception rays*. This is shown in figure 2.7.

At every time-step a ray is sent into space in a random direction. The randomness for the directions is shaped, so that the pdf with respect to the angle corresponds to the axiomatic uniformity of visual attention in Eq. (2.1). This is accomplished by three Gaussian functions [64]. The rays intersecting an object are counted and averaged in real-time using exponential averaging. Normalizing the amount of rays intersecting an object with the total amount of rays sent into the scene in the same time interval yields the degree of perception of the object. The exponential average P_q at time-step q of the perception process quantifies the perception of the corresponding object, and it is obtained by

$$P_q = \omega P_{q-1} + (1-\omega) p_q , \tag{2.14}$$

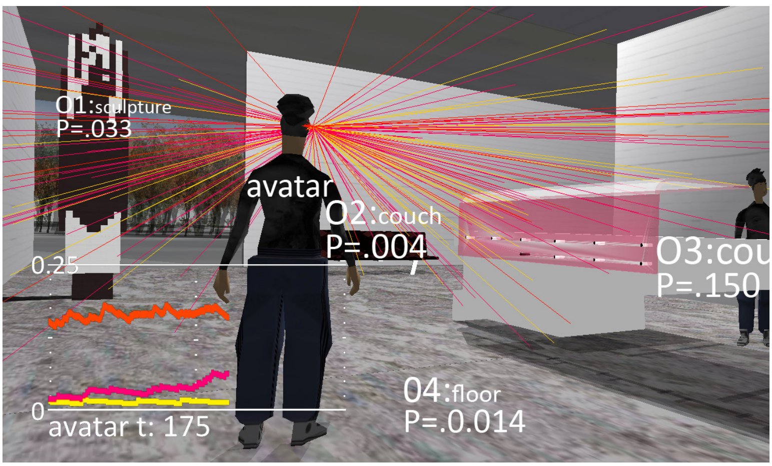


Fig. 2.7 Perception measurement in a scene in virtual reality

where p_q denotes a new perception signal to be integrated into the averaging process that stems from the interaction of a perception ray with the object; P_{q-1} denotes the old average at the previous time-step q-1; ω is a weight factor that is used to make a weighted summation between the old average and the new signal. The weight factor ω is obtained via a time constant τ as given by

$$\omega \equiv 1 - \frac{1}{\tau} \cdot \tag{2.15}$$

It is noted that the exponential average is based on the principle that the older a data sample is, which is integrated into the average, the less significance the sample has. The variation of the significance is given by an exponential function. It is important to realize that the summation of the elemental weights is equivalent to the real average of the data samples without weighting them.

In the averaging process, a ray reaching an object gives unity as input to the averaging process, while a ray hitting another object gives zero for the object concerned. This entails that for a ray hitting the object Eq. (2.14) becomes

$$P_q = P_{q-1} + \frac{1 - P_{q-1}}{\tau} \cdot \tag{2.16}$$

This means that the previous awareness of the object is increased by a value proportional to the difference between current object awareness and the total awareness for the scene. The proportionality factor is 1/ τ . In case the ray doesn’t hit the object Eq. (2.14) becomes

$$P_q = P_{q-1} - \frac{P_{q-1}}{\tau} \cdot \tag{2.17}$$

This means the awareness of the object concerned is reduced by a value proportional to the current object awareness that is scaled by the time constant τ. After sufficient processing time steps the resulting number P_q gives the degree of awareness of the object.

It should be noted that the time constant involved in the averaging process influences the accuracy of the obtained perception. A higher time constant yields a more accurate outcome at the cost of more time needed to obtain the result, compared to the case using a smaller time constant.

From Figure 2.7 it is clearly observed that the amount of perception rays hitting the different objects per unit time is different. For instance the counter labelled *O3* receives more perception rays compared to the couch positioned in the back of the room. This is due to the difference in size of the objects, and their relative position and orientation with respect to the avatar. This means that an unbiased observer is more aware of the counter than the couch. The results from the perception approximation are indicated in the figure. In this example .150 is the perception for the counter versus .004 for the couch. This means an unbiased observer is 38 times more aware of the counter than the couch when assuming the viewpoint of the avatar.

2.4 Implications of the theory for architectural design

2.4.1. Perception with varying observer location

A relevant question in architecture is how the unbiased perception differs when an observer assumes different viewpoints in the same spatial environment. This is a relevant issue for

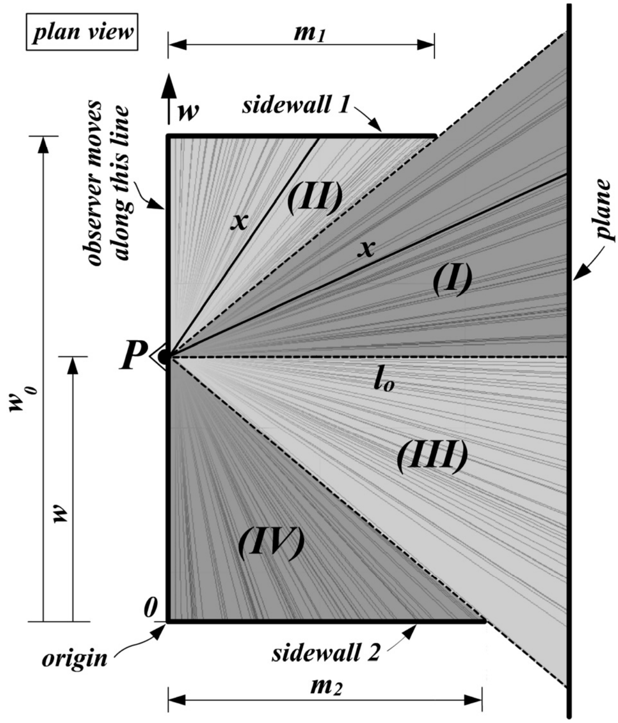


Fig. 2.8 Plan view of perception in a space; the observer moves along the left most wall-side, where the position *(w)* of an observer *(P)* is also indicated

example when deciding about the position of an entrance to a space, the location of a street entering an urban square, or the experience of an observer moving through an environment. To address this issue perception in a general spatial situation is computed, as shown in figure 2.8.

For any point within the semi-enclosure shown in the figure, the perception can be calculated using the probabilistic perception theory. First we compute the integral perception from the position of observation, indicated as *P* in the figure. Thereafter, we find the visual perception probability density with respect to the distance *w* by differentiation. In this way the variation of perception from the different positions along the leftmost wall-side is computed [8]. To accomplish the computation the whole region of interest is divided into four probabilistic regions. Each of the regions is considered separately. The results are eventually combined for the final outcome. The total probability density of perception for the given geometry is given by [8].

$$f_w(w) = \frac{1}{arctg\frac{w_o}{m_1} + arctg\frac{w_o}{m_2}} \left(\frac{m_1}{m_1^2 + (w_o - w)^2} + \frac{m_2}{m_1^2 + w^2} \right) \tag{2.18}$$

This is the visual attention along the line *w*. The associated perception is obtained by integrating $f_w(w)$ along an interval of *w*, where it incorporates all possible views to the environment from the region with respect to θ. This means this is an integral value with respect to θ and with respect to *w.*

It is interesting to note that although Eq. (2.18) is derived from the premise in Eq. 2.1 the parameter l_o, which is the distance to the frontal plane, does not play a role in it. This can be

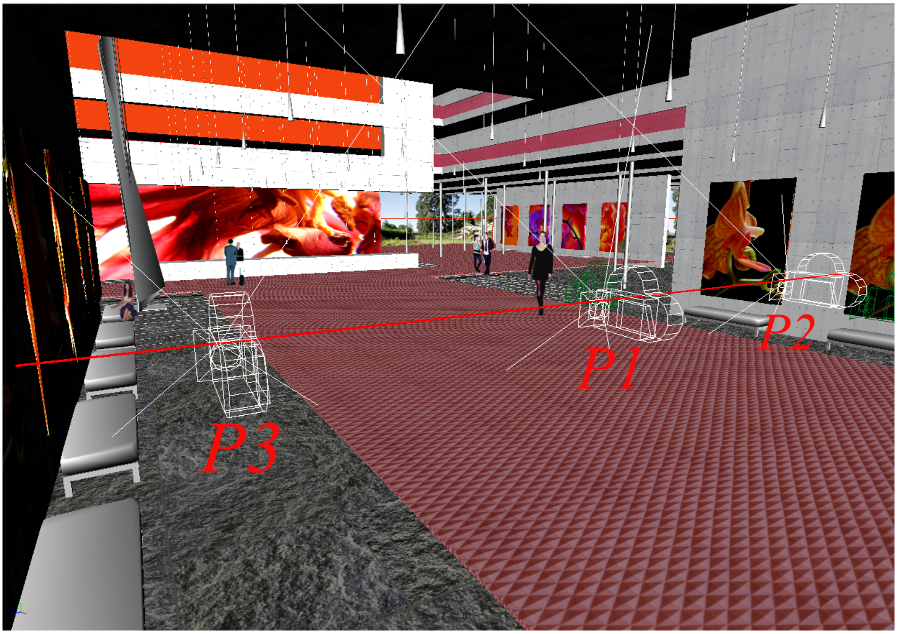


Fig. 2.9 Three separate observation positions *P1*, *P2*, *P3* used for the demonstrations

intuitively explained by considering that the plane is equivalently influencing the perception from all viewpoints along *w.* Therefore l_o does not affect the probability density from viewpoint to viewpoint.

2.4.2. Computer experiments on varying observer location

Based on the theoretical computations in the previous section, the variation of the visual perception along a horizontal line in space is demonstrated using several computer experiments. The line *w* in the space and three exemplary viewpoints on it are shown in figure 2.9. These viewpoints are used in figures 2.10, 2.11a and 2.11b. We note that figures 2.12 and 2.13 are taken from two from different viewpoints in another space.

Each figure shows a ray trace rendering of the scene, a plan view of perception situation, and the degree of perception $f_w(w)$ along *w*. In the plan view the respective viewing position is indicated with a letter '*P.'* The degree of perception belonging to this is pointed out with a red vertical line in the $f_w(w)$ graph.

To verify the results let us examine the scenes shown as follows. The scene depicted in the figures is a lobby. It contains a desk at the distance, several couches on the left, images on the walls, and a woman walking towards our viewing location. The ability of an unbiased observer to appreciate the visual information of the scenes varies from scene to scene. In other words, our degree of awareness of the visual content of the scenes is higher or lower depending on our viewpoint in the space.

The computations following from the probabilistic perception theory show that the perception demonstrated in figure 2.10 is minimal, figure 2.11a is high and figure 2.11b is low. This coincides with our common perception experience of the scenes. In figure 2.10 the distances to the walls are maximal due to our central viewing position. Therefore it is most difficult to obtain visual details from the walls. In figure 2.11a we obtain the details along the longer wall nearby our viewing location. This increases our perception.

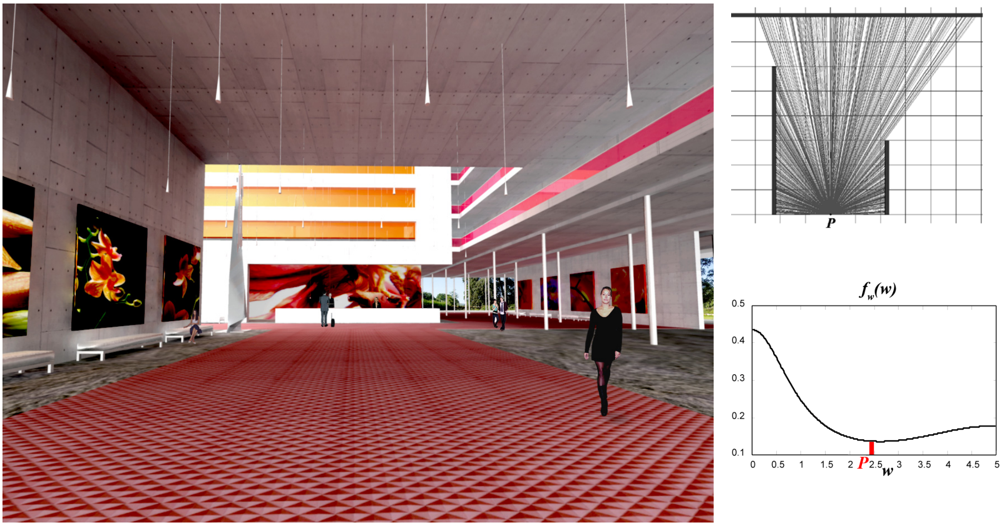
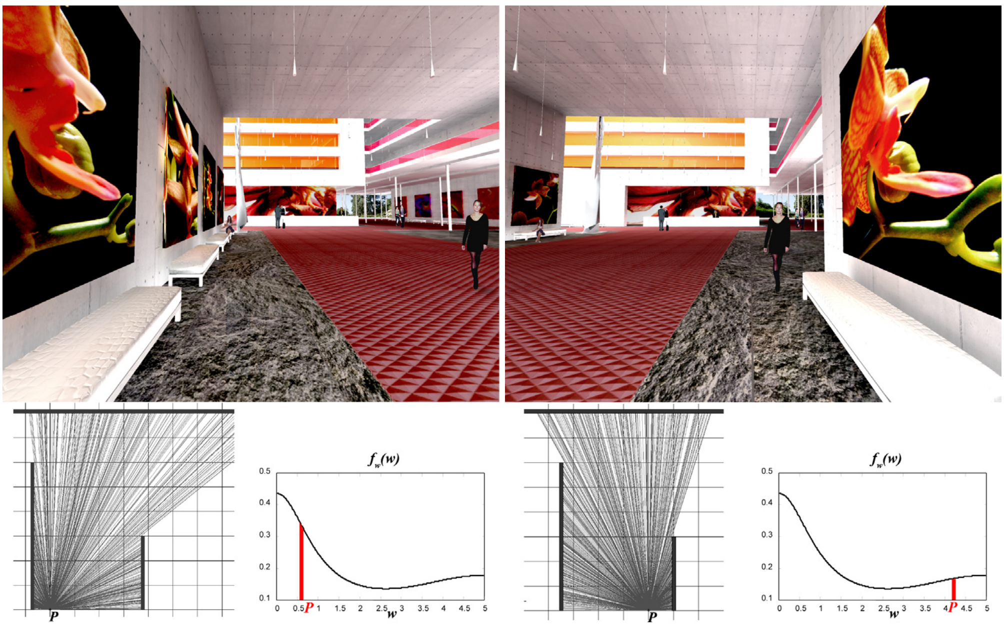


Fig. 2.10 Central viewing position *P1* with minimal degree of integral perception

Additionally a portion of the space in the far right side of the space is visible, which is hardly noticeable from the other viewpoints. In figure 2.11b we appreciate the details of the shorter wall nearby, which increases perception slightly compared to figure 2.10. These verifications are complemented by the perception demonstrations in figures 2.12 and 2.13, where the situation is contrary to the earlier example.



(a) (b)

Fig. 2.11 Left-hand viewing position P3 where the integral perception is high (a); right-hand viewing position P2, where the integral perception is low (b);

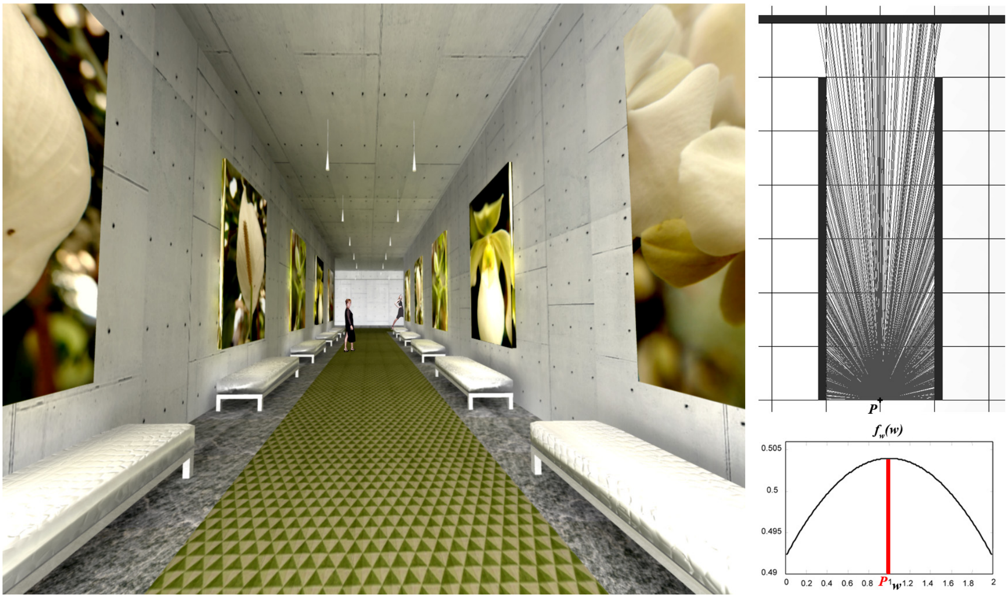


Fig. 2.12 Central viewing position where the degree of integral perception is maximal

The scene shown in these figures is a corridor-like space. The space contains couches on either side, images on the walls, and a woman standing at a distance. In this case the maximal and minimal perception positions are located at different places; namely the maximal perception occurs at the central position. This is shown in figure 2.12. The perception degree is less nearby the side walls. This is shown in figure 2.13. In contrast to the perception results found in the lobby space, in a narrow space the central position is advantageous for perception, as the computations revealed.

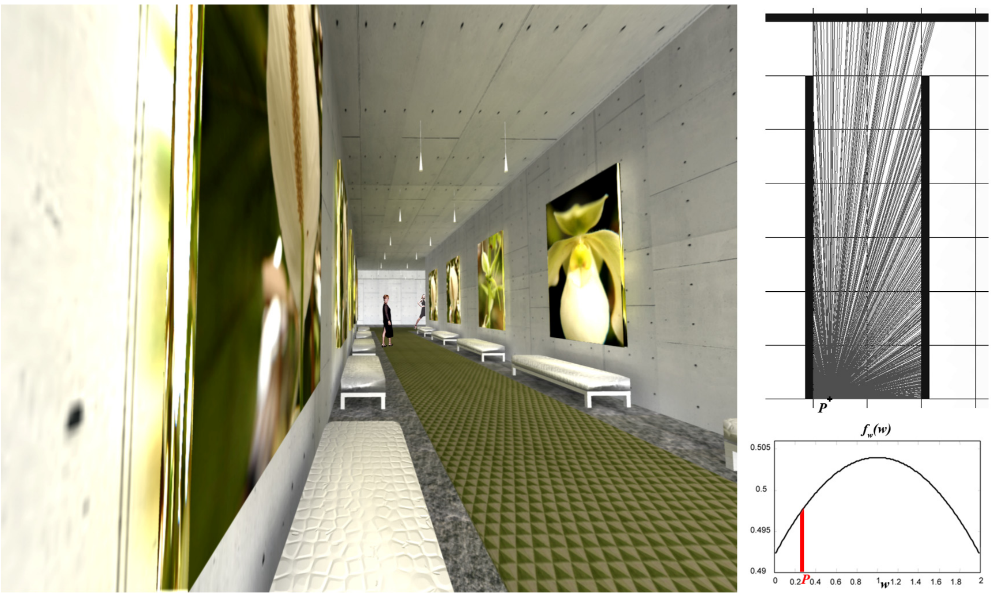


Fig. 2.13 Decentred viewing position, where the degree of integral perception is low

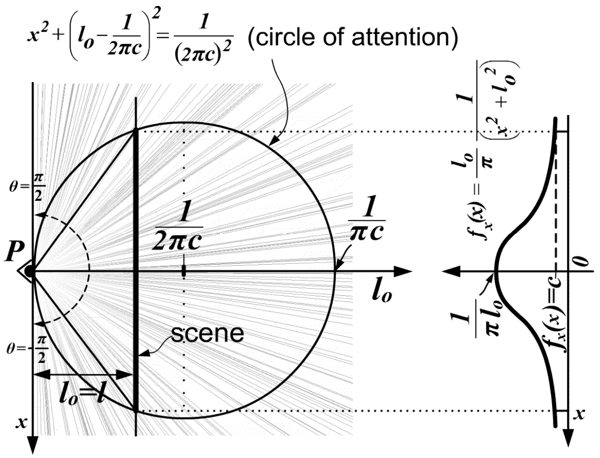


Fig. 2.14 Illustration showing *circle of attention,* which represents the probability density function of early visual perception, while the probability per unit length is taken as constant value *c*.

This can be qualitatively explained as follows. The more vision rays are inclined to be orthogonal to the side-walls, the more perception is obtained. Nearby the walls the majority of vision rays of the observer plunge into the space ahead. These results coincide with our intuitive judgment as to perception, which indicates the validity of the probabilistic perception theory and its applicability to architecture.

2.4.3. Influence of distance on the perception of scenes

The probabilistic perception theory can be used to compare architectural spaces with respect to perception defined above in mathematical terms. In particular the influence the distance between an observer and a scene has on perception is investigated below. This is a relevant issue when the basic dimensions of a space are determined. Depending on the intended purpose of a space an architect may for example wish to confront an observer entering the space with a high amount of visual information. Or he/she may intend to keep the visual perception degree purposely low. In an unbiased vision case this degree of perception is depending on the distance he/she has from the scene.

Let us compare scenes that are located at different distances. To make a 'fair' comparison we need to establish a common ground in the comparison. This is accomplished by considering that the visual attention viewing an infinite plane is a constant value. That is, the probability density $f_x(x)$ in Eq. (2.7) takes a constant value denoted by *c*. Its unit is probability per unit length. Taking the visual scope as $-\pi/2 \le \theta \le +\pi/2$, from Eq. (2.7) we write

$$\frac{1}{\pi}\frac{l_o}{x^2+l_o^2}=c\,. \tag{2.19}$$

This yields

$$x^2+(l_o-\frac{1}{2\pi c})^2=\frac{1}{(2\pi c)^2}\,. \tag{2.20}$$

Taking both *x* and l_o as variables Eq. (2.20) represents a circle in a coordinate system formed by *x* and l_o axis. The origin is located at the position of the observer. This is illustrated in figure 2.14, where position of the observer is denoted by *P*.

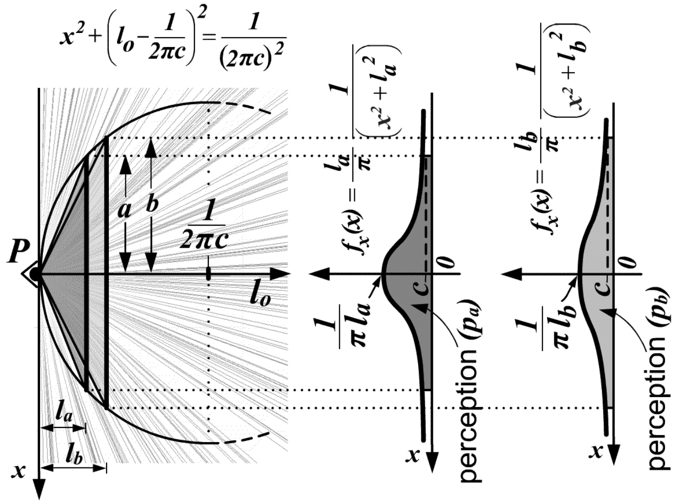


Fig. 2.15 Illustration of the effect of changing the viewing distance from l_a to l_b with respect to the viewed scene.

The corresponding early attention probability density is also shown in figure 2.14, where the attention threshold $f_x(x)=c$ is indicated. This circle is termed as *circle of attention*. Along the circle circumference the attention has a constant value; namely c. The inner part of the circle corresponds to a region of the space, where the unbiased degree of attention is deemed to be significant. Outside of the circle it is deemed insignificant. This is because outside of the circle the attention is $f_x(x)<c$. In this sense c is an attention threshold level, below which perception is considered not to occur. The attention threshold concept allows comparing scenes consistently when their distance to an observer is differing.

Let us compare the vision probability of a scene viewed from different distances. The constant attention threshold c is taken to be the same for both cases. This yields the same circle of attention in both cases. This is shown in figure 2.15, where the scenes in question are represented as bold vertical lines within the circle of attention seen in the figure.

For the sake of clarity of the explanation, first the scene in question is considered as a one-dimensional entity, extending in x-direction only. The distances between the viewpoint and the scenes are denoted as l_a and l_b. Clearly, $l_a<l_b$. The circle of attention intersects the scenes at different locations, so that the portions of the scene within the circle have different size. The portions are denoted as a and b in figure 2.15.

The unbiased perceptions of the scenes differ depending on the distance the scene is viewed from. In figure 2.15 the respective probability density functions $f_x(x)$ that belong to the different distances are sketched. The vision probabilities p_a and p_b respectively in the cases of $l_o=l_a$ and $l_o=l_b$ are indicated by the shaded areas in the $f_x(x)$ vs x plots. The probabilities correspond to the viewed scenes, which are shown within the circle of attention. To obtain the difference of the scenes with respect to their respective vision probabilities p_a and p_b, the latter are computed as follows.

For $l_o=l_a$, from figure 2.15, the integral of attention, as perception, yields

$$p_a=\int_{-a}^{+a} f_x(x)\,dx=\frac{l_a}{\pi}\int_{-a}^{+a}\frac{dx}{x^2+l_a^2}=\frac{2}{\pi}arctg\frac{a}{l_a} \qquad (2.21)$$

For $l_o=l_b$, from figure 2.15, the integral of attention yields

$$p_b = \int_{-b}^{+b} f_x(x)\,dx = \frac{l_b}{\pi}\int_{-b}^{+b}\frac{dx}{x^2+l_b^2} = \frac{2}{\pi}arctg\frac{b}{l_b} \tag{2.22}$$

As an illustrative example, assuming that *a=0.65 R* where *R* is the radius of the circle of attention, l_a is computed from figure 2.15 as

$$l_a = R - \sqrt{R^2 - a^2} \cong 0.24R \tag{2.23}$$

so that from Eq. (2.21) the perception p_a becomes

$$p_a \cong \frac{2}{\pi}\text{arctg}\frac{0.65R}{0.24R} \cong \frac{2}{\pi}1.22 \cong 0.78 \tag{2.24}$$

Assuming that *b=0.8 R*, l_b is computed from figure 2.15 as

$$l_b = R - \sqrt{R^2 - b^2} = 0.4R \tag{2.25}$$

so that from Eq. (2.22) the perception p_b becomes

$$p_b = \frac{2}{\pi}\text{arctg}\frac{0.8R}{0.4R} \cong \frac{2}{\pi}1.11 \cong 0.70 \tag{2.26}$$

Summarizing the computations, for the geometry in figure 2.15, the vision probability is higher for near distance l_a relative to the case with far distance l_b. In the figure the geometrical proportions are given by

$$\frac{a}{l_a} \cong 2.71 \text{ and } \frac{b}{l_b} = 2.00 \tag{2.27}$$

$$\frac{l_a}{l_b} \cong \frac{0.24R}{0.40R} \cong 0.60 \tag{2.28}$$

and the respective perceptions are given by

$$p_a \cong 0.78 \text{ and } p_b \cong 0.70 \tag{2.29}$$

The result in Eq. (2.29) shows that the perception p_a is more than p_b.

2.4.4. Two-dimensional case for perception of scenes

The difference in perception between the two scenes is accentuated by the fact that perception of the scene is more accurately modelled as a two-dimensional event. This is in contrast to the computation above, where the perception is modelled as one-dimensional event. This is accomplished as follows.

In the two-dimensional case perception is considered in horizontal direction and vertical direction. For the horizontal direction the probability density along the *x* axis is considered. The axis is shown in figure 2.1. Taking the visual scope as $-\pi/2 \le \theta \le \pi/2$ the corresponding attention probability density is [8]

$$f_x(x) = \frac{1}{\pi}\frac{l_o}{x^2 + l_o{}^2} \tag{2.30}$$

and an elemental perception during early vision is

$$P_x = \int_{\Delta x} f_x(x)\,dx \cong f_x(x)\,\Delta x \tag{2.31}$$

Fig. 2.16 Scene corresponding to the perception computation shown in figure 2.15 where $l_o=l_a$=6m; the degree of perception of the scene is $P_{x,z}\cong 0.16$

For the vertical direction the same geometry shown in figure 2.1 applies also, but next to the *x*-axis, perception is considered along another axis. The new axis is *z*-axis shown in figure 2.1. The perception along the *z*-direction is taken with the same visual scope as in *x* direction, so that the corresponding attention probability is

$$f_z(z) = \frac{1}{\pi}\frac{l_o}{z^2 + l_o^{\;2}} \tag{2.32}$$

and an elemental perception is

$$P_z = \int_{\Delta z} f_z(z)\,dz \cong f_z(z)\,\Delta z \tag{2.33}$$

It is important to note that in the two-dimensional perception case the events of perceiving an object in *z*-direction vs. *x*-direction are *independent events*. Therefore, the joint probability of the perception is computed by means of multiplication of the vision probabilities belonging to each dimension. This is given by Eqs. (2.30) and (2.32). Explicitly we write

$$P_{x,z}(x,z) = P_x\,P_z = \int_{\Delta x}\frac{1}{\pi}\frac{l_o}{x^2 + l_o^{\;2}}\,dx \;\int_{\Delta z}\frac{1}{\pi}\frac{l_o}{z^2 + l_o^{\;2}}\,dz \tag{2.34}$$

The effect of distance on early perception is exemplified in two scenes given in figure 2.16 and figure 2.17. Each scene corresponds to the same circle of attention shown in figure 2.16. In the figures the entrance hall of a building is shown. A group of persons and several objects can be seen that form a scene, corresponding to the scenes schematically shown in figure 2.15.

Fig. 2.17 Scene corresponding to the perception computation shown in figure 2.15 where l_o=l_b=10m; the perception of the scene is $P_{x,z}$≅*0.09.*

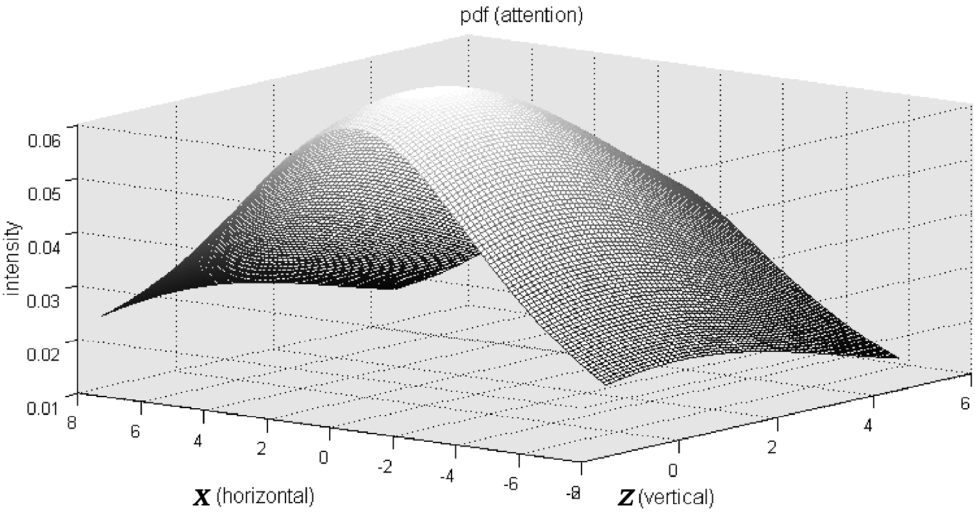
In figure 2.16 the scene is viewed from a near distance corresponding to l_a in figure 2.15. In figure 2.17 the same scene is viewed from a far distance corresponding l_b in figure 2.15. It is noted that in Eq. (2.34) *dx* and *dz* are infinitesimally small distance intervals around *x* and *z.* In the following perception comparison the joint probability given by Eq. (2.34) is approximated by

$$P_{x,z}(x,z) \cong \frac{1}{\pi^2} \frac{l_o^{\,2}}{(x^2 + l_o^{\,2})(z^2 + l_o^{\,2})} \Delta x \, \Delta z \tag{2.35}$$

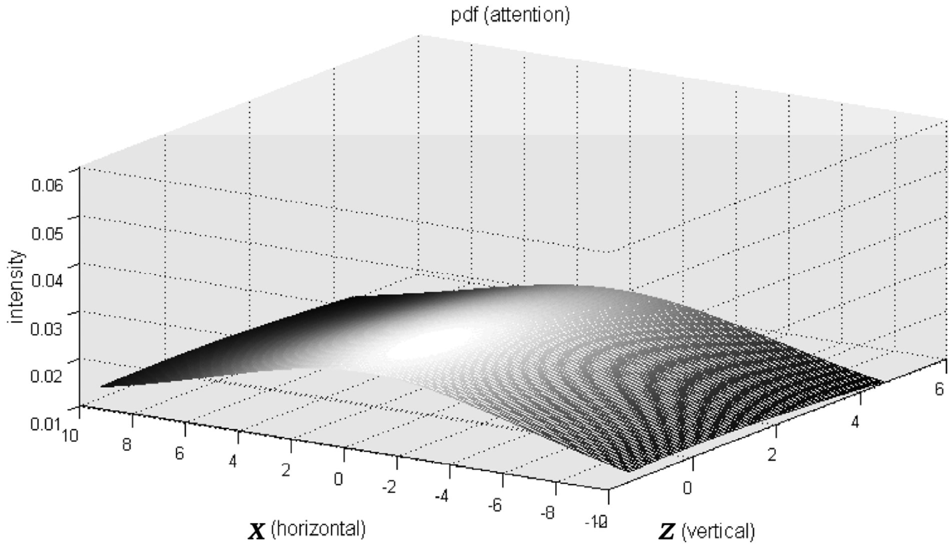
taking *Δx* and *Δz* as sufficiently small distances. In figures 2.16 and 2.17 the degrees of elemental early vision probabilities are given forming an array of numbers. The probabilities belong to patches with the size *ΔxΔz*. A degree is plotted at the centre of the respective patch. The summation of the elemental perceptions yields the resulting vision probability of the scene.

Viewing the scene as shown in figure 2.16 the resulting perception is $P_{x,z}$≅*0.16* while viewing as shown in figure 2.17 yields $P_{x,z}$≅*0.09* ; the parameters involved in the perception computations have the following values: *a*=16m, l_a=6m, *b*=20m, l_b=10m, and *c*=0.0064m^{-1}.

The scene has a height of 6m. Clearly the perception of the scene as viewed in figure 2.16 is higher than figure 2.17. This means that the probability of getting the visual information present in the scene shown in figure 2.16 is higher than for the scene shown in figure 2.17. The increase in perception is about factor 1.8 although the distance is only decreasing by about factor 1.6. This indicates the relevance of distance on perception during early vision. To illustrate the difference the plot of the two-dimensional perception $P_{x,z}(x,z)$ along the *x* axis (horizontal) and *z* axis (vertical) is shown in figure 2.18.



(a)



(b)

Fig. 2.18 Two-dimensional pdf (attention) that belong to the scenes shown in figures 2.16 and 2.17; plot (a) corresponds to figure 2.16; plot (b) corresponds to figure 2.17

The result obtained corroborates with our common perception experience: Viewing a scene from closer distance more visual information is obtained. Since the difference in terms of unbiased vision probability is quantified, spaces can be precisely compared in architectural design. An architect can determine the degree of visual detail of a scene that should come to the awareness of a person during early vision by defining the spatial envelope accordingly. We can determine this perception degree using the model presented above.

Occupants and clients of buildings have a general demand for large spaces, which have a lower degree of perception with respect to their spatial envelope. The demand is

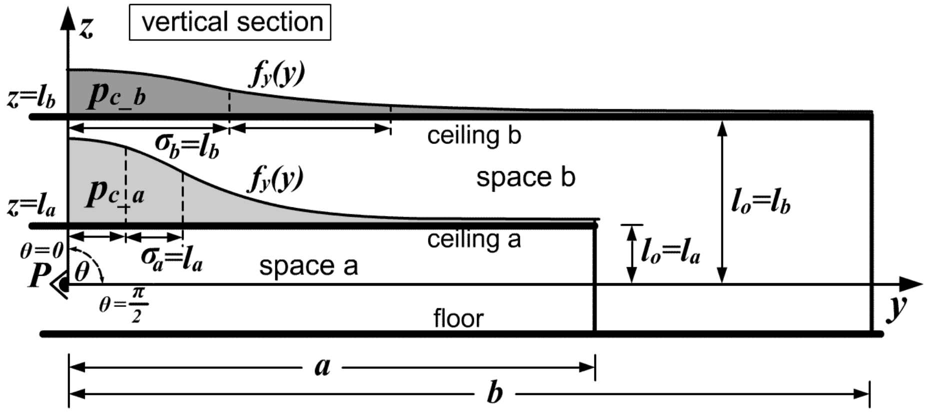


Fig. 2.19 Vertical section illustrating the effect of changing the ceiling height from l_a to l_b on the perception of the ceiling during early vision.

presumably due to the fact that high degree of perception of the spatial envelope is distracting during other mental activities. In other circumstances a high degree of perception of a scene is demanded, for example in case of a stadium or a concert hall. In any of these cases the positioning of the viewing locations and the spatial elements has to take the perceptual demand into account.

2.4.5. Influence of distance on the perception of ceilings

Another example showing the significance of visual perception for architecture concerns the choice of the ceiling height. Proportions of a space influence how strongly we are aware of the ceiling of the room. When our awareness of the ceiling is above a certain degree this may be undesirable depending on the intended purpose of the space. For example awareness of a ceiling may be distracting or cause a feeling of 'oppressiveness,' while this experience may conflict with the purpose of the space. For example in case of an entrance hall it is required that the space feels *open* yields an impression of *freedom* and *interaction*. In other cases awareness of the ceiling is demanded in order to create a feeling of spatial *intimacy,* for instance in waiting rooms, or to contribute to the richness of the ambience, as for example in Baroque churches. In the following the influence of ceiling height on perception in unbiased vision is exemplified. Figure 2.19 shows a section view of two spaces labelled *space a* and *space b*. The spaces have different ceiling heights l_a and l_b and also different lengths *a* and *b*.

The degree of perception of a ceiling differs depending on the distance the ceiling is viewed from and the size of the ceiling. In figure 2.19 the respective probability density functions $f_y(y)$ that belong to the different ceiling heights are sketched.

The unbiased vision probabilities of *ceiling a*, denoted by p_{c_a} and of *ceiling b*, denoted by p_{c_b} are indicated by the shaded areas in the $f_y(y)$ vs *y* plots. The characteristic distance l_o is the distance between ceiling and the eye height of the observer and denoted as $l_o=l_a$ and $l_o=l_b$ for the respective ceilings. To obtain the difference of the ceilings with respect to perception, p_{c_a} and p_{c_b} are computed as follows. For the sake of clarity of the explanation first perception along the *y*-axis is considered, modelling the perception along one spatial dimension.

We take the scope of vision as $0 \leq \theta \leq \pi/2$, which yields

$$f_y(y) = \frac{2}{\pi} \frac{l_o}{(l_o^2 + y^2)} \quad (0 \leq y \leq \infty) \tag{2.36}$$

as probability density function of perception. For $l_o = l_a$, from figure 2.19, the integral of attention yields

$$p_{c\_a} = \int_0^a f_y(y)\,dy = \frac{2l_a}{\pi} \int_0^a \frac{dy}{y^2 + l_a^2} = \frac{2}{\pi} arctg \frac{a}{l_a} \tag{2.37}$$

For $l_o = l_b$, from figure 2.19, the integral of attention yields

$$p_{c\_b} = \int_0^b f_y(y)\,dy = \frac{2l_b}{\pi} \int_0^b \frac{dy}{y^2 + l_b^2} = \frac{2}{\pi} arctg \frac{b}{l_b} \tag{2.38}$$

As an illustrative example, assuming that a/l_a=10 the early vision probability p_{c_a} becomes

$$\mathrm{p}_{c\_a} = \frac{2}{\pi} \mathrm{arctg} 10 \cong \frac{2}{\pi} 1.47 \cong 0.94 \tag{2.39}$$

Assuming b/l_b=5 the perception p_{c_b} becomes

$$\mathrm{p}_{c\_b} = \frac{2}{\pi} \mathrm{arctg}\, 5 \cong \frac{2}{\pi} 1.37 \cong 0.87 \tag{2.40}$$

The results in Eqs. (2.39) and (2.40) show that the perception of *ceiling a*, denoted by p_{c_a} is higher than the perception of *ceiling b*, denoted by p_{c_b}.

2.4.6. Two-dimensional case for perception of ceilings

The difference in perception between the two ceilings is accentuated by the fact that the perception is more accurately modelled as a two-dimensional event, in contrast to the computation above, where the perception is considered along one spatial dimension. This is accomplished as follows.

In the two-dimensional case unbiased perception is considered in forward direction and horizontal direction. For the forward direction the probability density along the *y* axis is given by Eq. (2.36). An elemental unbiased perception is given by

$$P_y = \int_{\Delta y} f_y(y)\,dy \cong f_y(y)\,\Delta y \tag{2.41}$$

For the horizontal direction the same geometry shown in figure 2.19 applies also, and next to the *y*-axis we consider perception along the *x*-axis shown in figure 2.1. As visual scope along the *x*-axis $-\pi/2 \leq \theta \leq \pi/2$ is taken. This way the corresponding attention becomes

$$f_x(x) = \frac{1}{\pi} \frac{l_o}{x^2 + l_o^2} \tag{2.42}$$

and an elemental perception during early vision is

$$P_x = \int_{\Delta x} f_x(x)\,dx \cong f_x(x)\,\Delta x \quad \cdot \tag{2.43}$$

It is important to note that in the two-dimensional perception case the events of perceiving an object in *y*-direction vs. *x*-direction are *independent events*. Therefore, the joint probability of the vision probabilities is computed by means of multiplication of the probabilities belonging to each dimension, given by Eqs. (2.36) and (2.42). Explicitly we write

$$P_{x,y}(x,y) = P_x P_y = \int_{\Delta x} \frac{1}{\pi} \frac{l_o}{x^2 + l_o^{\,2}} dx \int_{\Delta y} \frac{2}{\pi} \frac{l_o}{y^2 + l_o^{\,2}} dy \tag{2.44}$$

We note that in Eq. (2.44) *dx* and *dy* are infinitesimally small distance intervals around *x* and *y.* In the following perception comparison the joint probability given by Eq. (2.44) is approximated by

$$P_{x,y}(x,y) \cong \frac{2}{\pi^2} \frac{l_o^{\,2}}{(x^2 + l_o^{\,2})(y^2 + l_o^{\,2})} \Delta x \Delta y \tag{2.45}$$

taking *Δx* and *Δy* as sufficiently small distances.
The effect of ceiling height on the perception of a ceiling during early vision is exemplified in two scenes given in figure 2.20 and figure 2.21. Figure 2.20 shows a space corresponding to *space a* in figure 2.19, where a=15m and l_a=1.5m. Figure 2.21 shows a space corresponding to *space b* in figure 2.19, where b=22m and l_b=4.4m. In both cases the ceiling is 18m wide, i.e. it extends 18m in *x*-direction.

In figures 2.20 and 2.21 the degrees of elemental perceptions are given as an array of numbers. Each number gives the early vision probability belonging to a ceiling patch with the size *ΔxΔy,* and the probability is plotted at the centre point of the respective patch. The summation of these elemental perceptions yields the resulting degree of perception of the ceiling.

The perception degree of the ceiling shown in figure 2.20 is $P_{x,y} \cong 0.94$, while it is $P_{x,y} \cong 0.50$ for the ceiling shown in figure 2.21*.* It is important to note that the probabilities computed

Fig. 2.20 Space corresponding to the perception computation shown in figure 2.19, where l_o= l_a=1.5m; the resulting degree of perception of the ceiling is $P_{x,y} \cong 0.94$.

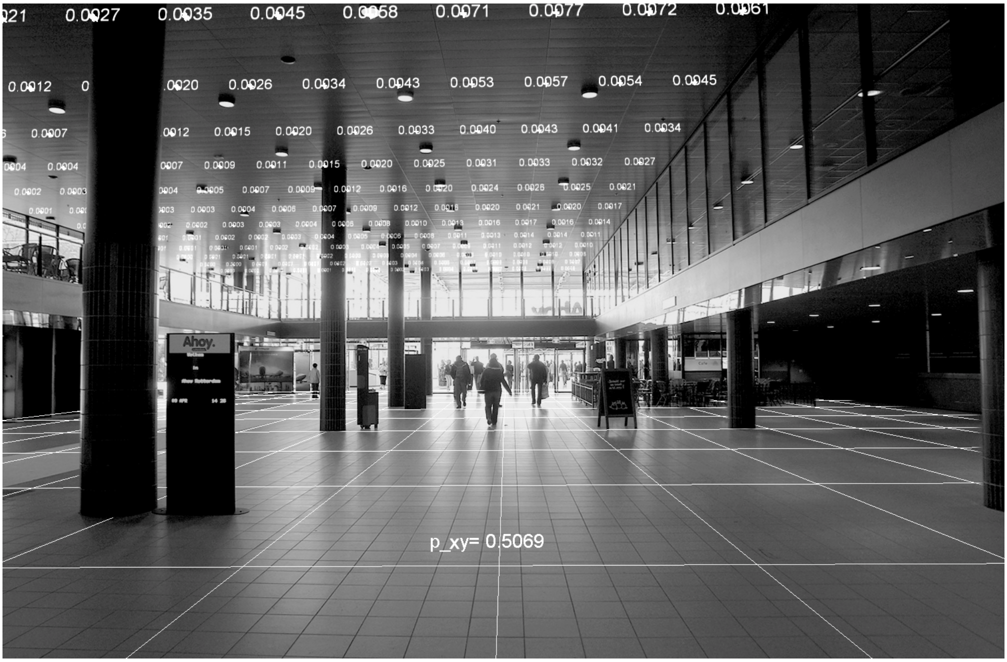


Fig. 2.21 Space corresponding to the perception computation shown in figure 2.19, where l_o= l_b=4.4m; the resulting degree of perception of the ceiling is $P_{x,y} \cong 0.51$.

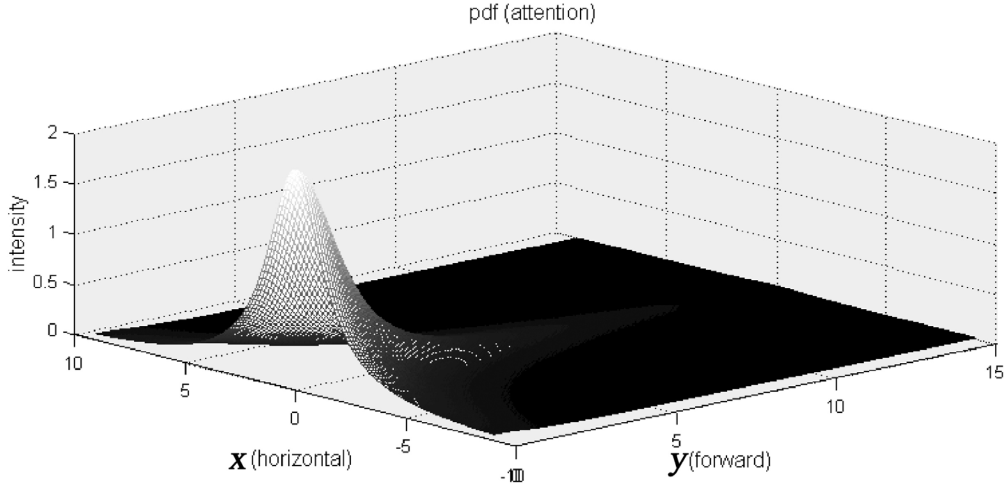
in the present case belong exclusively to the ceiling and not the entire space. This is done for appropriate comparison of the perceptions. Although the ceiling of the space shown in figure 2.20 is smaller in size compared to the ceiling shown in figure 2.21, the former has a significantly higher degree of perception compared to the ceiling of the latter space. The degree of perception of the ceiling in figure 2.20 is about factor 1.8 higher compared to the ceiling in figure 2.21. The meaning is that the ceiling shown in figure 2.20 has almost twice the visual impact compared to the ceiling shown in figure 2.21. To illustrate the difference the plot of the two-dimensional perception $P_{x,y}(x,y)$ along the axis x (horizontal) and y (forward) is shown in figure 2.22. In the figure the location of the observation point is at $x=y=0$.

From figure 2.22 we note that the major contribution to perception during unbiased vision stems from the region of ceiling nearby the observation point. Here the visual attention is relatively high compared to distant regions.

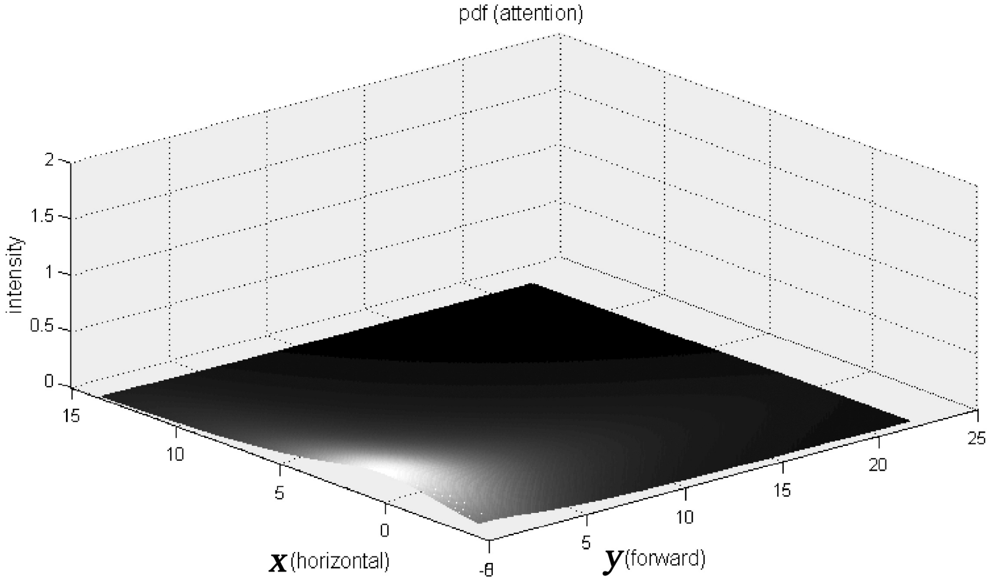
The perception of ceiling is an important issue in architecture as ceilings are most common spatial elements. Often it is undesirable that the ceiling has a high degree of perception. This is because generally the ceiling is not an environmental object to carry

relevant visual information, but instead is a necessary element for other reasons. A ceiling with a high degree of perception or attention can be considered disturbing as it distracts from more relevant information in the environment. This issue becomes particularly noticeable in spaces with large dimensions, where the ceiling will have a substantial degree of perception unless it is placed at sufficient height. Architects know this by experience; however it was not possible to quantify the perceptual effect precisely and to compare different situations on a common ground.

It is interesting to note that the attention pdf is well approximated by a Gaussian, where the standard deviation σ is found to be $\sigma \approx l_o$ [14]. This implies $\sigma_a \cong l_o$ and $\sigma_b \cong l_b$, which is



(a)



(b)

Fig. 2.22 Plots of the pdf (attention) that belong to the ceilings shown in figure 2.20 and 2.22; plot (a) corresponds to figure 2.20; plot (b) corresponds to figure 2.21

indicated in figure 2.19. As the statistical properties of the Gaussian are well known, the relevance of attention/perception in architectural design can be easily interpreted. Namely, having a distance l_o between viewpoint and ceiling, if we consider a region of the ceiling of length l_o in *y*-direction shown in figure 2.19, the early vision probability becomes about 0.68. In the same way, if we consider a region of length $2l_o$ the vision probability becomes about 0.95. In horizontal direction, if we consider a region of the length $2l_0$ in *y* direction the vision probability becomes 0.68.

As a rule of thumb for architects to calculate the vision probability of a ceiling we can consider a square shaped surface patch extending with the length $2l_o$ in *y*-direction and $2l_o$

in x-direction. The vision probability of the patch is approximately $P_{x,y} \cong 0.68 \times 0.95 \cong 0.65$. Considering a patch that extends with $2l_o$ in y-direction and $4l_o$ in x-direction, the perception of the patch is approximately $P_{x,y} \cong 0.95^2 \cong 0.90$.

It is emphasized that in these computations on the perception of ceilings the vision is taken within a scope spanning 90 degrees, from looking vertically upwards to looking horizontally. One may wonder how the computations would change in case the vision would span both floor and ceiling. In this case it is convenient to consider the visual scope in two parts, one part spanning the ceiling and one part spanning the floor. In the computations for perception of the floor the coordinate system in figure 2.19 is mirrored at the horizontal axis, so that the z axis points downwards, and as lo value in eq. (2.36) the distance between eye and floor applies. It is clear that integrating the respective pdfs along floor and ceiling and summing them up is to yield unity. Therefore the pdf expressed in eq. (2.36) should be multiplied by factor .5 to apply for both floor and ceiling. It is noted that although the shapes of the pdfs for floor and ceiling differ by the different distance l_o their integral from y=0 to y=infinity yields the same number, i.e. 0.5. It is also noted that to obtain the analytic form of the pdf of perception along any object is described in a publication [58].

2.4.7. Visual openness perception

Based on the definition of visual perception derivatives of perception can be also defined. One of such derivatives is *visual openness* perception. In this thesis visual openness is described as the visual perception of long distances. It is derived from the perceived distances between observer and environment. These are the distances from the environmental objects an observer is aware of.

We can consider distance in terms of long or short distance for instance. It is clear that the concept of large distance does not comprise a sharply defined threshold. That is, there is no particular distance from on which distances would be considered large, and before which they would not be. For instance it would be unnatural to term an interior wall at 5m distance to be *at a long distance,* while if it were 4,90m the distance would not be considered long anymore. From this example it is seen that classical set theory, where sets are sharply defined, is inappropriate to classify distances as long distance. Due to the fuzzy definition of the class boundary it is appealing to consider *long distance* as a fuzzy set rather than a crisp set. This means the degree of membership to the set of far distances can be expressed by means of a fuzzy membership function. The fuzzy membership function is selected as a sigmoid function given by

$$f(x) = \frac{1}{1+e^{-(x-x_o)}} \quad .$$

The variation of the sigmoid with the independent variable *x,* which represents the mean perceived distance, is shown in figure 2.23. The characteristic behaviour of the function is its saturation at the extremities and its approximately linear behaviour in the middle range.

Based on the definition of *long distance* as a fuzzy set, let us consider an environment that contains a mixture of different distances spanning between observer and the visible surfaces of the environment. For each elemental distance the observer perceives its membership to the set of long distance is identified using the sigmoid. When we compute the average of the membership degrees, this value characterizes the space from a perceptual viewpoint. This mean of the membership degrees for the perceived distances is defined as perception of *visual openness* [1].

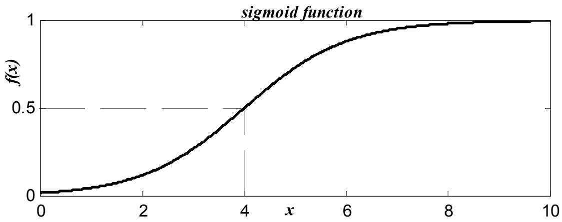


Fig. 2.23 The sigmoid function, $f(x)=1/[1+exp-(x-x_o)]$, with $x_0=4$

Qualitatively, for a space where the mean perceived distance is small, due to the sigmoid function as membership function the perception of the visual openness for this space is considered to be small, too, with no significant change in this fuzzy range. Meaning a space that is considered to have already low visual openness is not considered to have a significantly lower openness if it reduces even further in size.

A similar behaviour is observed at the other extreme, where the visual openness perception does not change significantly as the mean perceived distance approaches to extreme values. That is a large space, or a space with openings to the exterior, where mean perceived distances are large, will also yield a high openness. Also the openness degree does not significantly increase as mean perceived distance is further increasing. We can term the latter effect as *saturation* of the membership degree.

At the middle range, the visual openness perception is highly dependent on the mean perceived distance, as one should expect. This means two medium sized environments that differ notably with respect to their mean perceived distances will also differ notably with respect to their openness perception.

These qualitative observations about the openness perception conform to our common, everyday visual openness perception. It is noted that the openness model proposed is similar to many other biological processes of human beings. An example is certain neuronal behaviour in the brain, which is also modelled using a sigmoid.

It is emphasized that due to the use of the sigmoid the perception is measured quantitatively in the range between 0 and 1. This is significant, especially since the virtual human is experiencing and evaluating the visual openness of a space as a fuzzy statement. It is noted that other shapes for the membership functions are considerable and subject to experimental verification. Empiric identification of optimal membership function parameters for openness measurement of interior spaces, and validation of this model is found in the literature [2, 4].

2.5 The influence of colour in perception

The present section focuses on the role colour plays during unbiased vision. It is common experience that an object having a similar colour as its background is less likely to be noticed than an object that has a colour which is strongly differing from the background colour. This is illustrated in figure 2.23. In the extreme case, if the colour of an object is the same as its background it is not perceived at all.

2.5.1. Colour and colour difference

Colour is an important feature to be included in the perception model, since it is fundamentally responsible for the perception phenomenon. Colour difference is necessary,

(a) (b)

Fig. 2.24 Effect of colour difference on perception of a building component

so that we see the object at all. That is, without any colour difference we do not distinguish an object from its background.

Colour is a perceptual effect of light. This means colour exists in human consciousness and not “in nature”. When we say an object *has* a colour the colour we attribute to the object is the content we have in our consciousness. This content we have in our consciousness due to sensory and neural processing. The processing is caused by a certain portion of the light spectrum the object reflects into our eye.

From figure 2.24 we note that the red building element in figure 2.25b is more easily noticed than the grey one in 2.25a. In other words our *degree of awareness* of the right object is higher compared to the left one.

The purpose of this section is to quantify the effect of increased/reduced awareness due to colour difference, and to integrate it into the geometry based perception model presented in the previous part of this chapter. In order to quantify the effect it is necessary first to specify a colour mathematically and then compute colour difference.

Colour difference expresses the difference between two colours. Informally it is also referred to as colour contrast. To obtain colour difference quantitatively it is necessary to specify the colours numerically. That means a certain colour is expressed as a point in a colour space. This is explained in sections A.1-A.3 of Appendix A. Having specified two colours as two points in a colour space, such as *CIE XYZ* or *CIE RGB space,* intuitively we may consider that the Euclidian distance measured between the two points indicates the difference between the colours. However for this to be the case it is necessary that the colour space obeys a certain criterion. This criterion is known as *perceptual uniformity*.

A colour space is perceptually uniform when the colour difference a person with normal vision experiences between any two colours specified in this space is equal to the Euclidian distance measured between the colours’ coordinates. Based on the C.I.E. standard observer two colour spaces that are approximately perceptually uniform are the C.I.E. $L^*u^*v^*$-Space and *CIE 1976* $L^*a^*b^*$-Space [65].

Having specified two colours in $L^*u^*v^*$ or $L^*a^*b^*$ space the colour difference ΔE^* between the colour is computed as the Euclidian distance between their chromaticity coordinates. This is given by Eq. (2.46) *for* $L^*u^*v^*$ and by (2.47) for $L^*a^*b^*$ space*.* These equations are knows as *the colour difference formulae.*

$$\Delta E^*_{uv} = \sqrt{(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2} \tag{2.46}$$

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \tag{2.47}$$

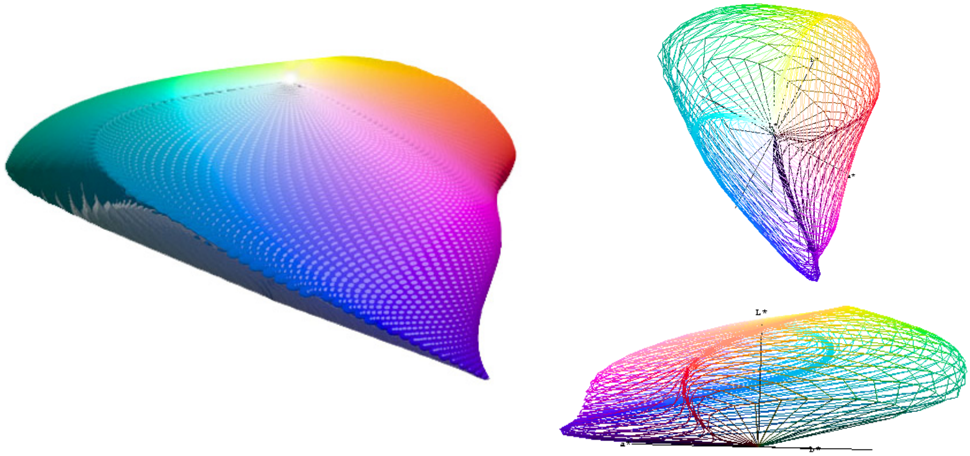


Fig. 2.25 The *CIE* $L^*a^*b^*$ space

It is noted that the *L** function defines the lightness correlate in both *CIE L*u*v** and *L*a*b** spaces. No simple relation exists between the chromaticity coordinates *u*, v** and *a*, b**.

In order to illustrate the non-linearity of perceptual uniformity, the shape of the L'a'b' colour space is shown in figure 2.25. From the figure we note that based on *the C.I.E. standard observer* the colour difference between black and white is significantly less than for example between a saturated purple and yellow colour. This is seen from the figure as follows. The line passing vertically through the space labelled *L** has black and white colour at its respective extreme points, and the length of this line is clearly less than a line connecting a purple point with a yellow one passing diagonally through the shape.

A second significant value for our purposes of extending the perception model with colour component is the *mean colour difference* among the elementary colours. This value is given by Eq. (2.48). The role this value plays in our perception model is explained below.

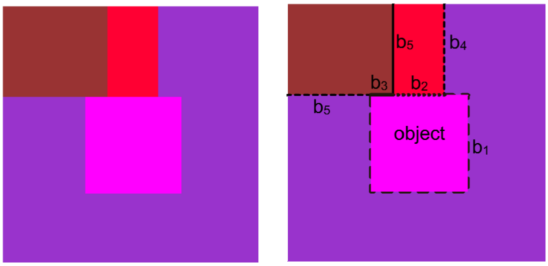
$$\sum_{n=1}^{n=28} \Delta E_n^* \approx 2638 \text{, so that } \overline{\Delta E}_n^* \approx 2638/28 = 94 \tag{2.48}$$

2.5.2. Colour difference and perception

The perception of an object is influenced by the colour difference between the object and its background. This component of perception is termed *colour perception* and denoted by P_c. In this work the colour perception is given by

$$P_c = \sum_{i=1}^{i=j} w_i \frac{\Delta E_i^*}{\overline{\Delta E^*}} = \frac{1}{\overline{\Delta E^*}} \sum w_i \, \Delta E_i^* \tag{2.49}$$

where ΔE_i^* gives the colour difference at the edge portion *i* of the object; $\overline{\Delta E^*}$ is the mean colour difference with respect to all edges delimiting the objects forming the scene in question weighted by the factor w_i. It is noted that for general comparison purposes the value given by Eq. 2.48 may be used as mean colour difference in Eq. (2.49). The value *j* gives the total amount of edges delimiting the object. The factor w_i in Eq. (2.50) is given by



(a) (b)

Fig. 2.26 A colour patch that is delimited with respect to 3 different neighbouring colour patches (a); indication of the borderline portions (b)

$$w_i = \frac{b_i}{b_1 + b_2 + \ldots + b_n}\ , \qquad (2.50)$$

where b_i is the length of the edge *i* in question. This means w_i gives the proportional length of a borderline segment b_i with respect to the sum of all borderlines of the scene. An example is shown in figure 2.26. It is emphasized that the length of a borderline is normalized using the sum of *all* borderline-lengths of the scene. This way proportions of an object play a role, beyond the object's size in terms if its surface area. For example an object with a compact shape, such as a cube, compared to an oblong shaped object has a shorter borderline with the background relative to the perceptually notable surface of the object. Therefore the boundaries of a compact shaped object will have a smaller w_i value compared to an object with an elongated shape. It is noted that the colour perception component is quantified by a probability. This probability belongs to the event that an edge is seen by the observer. The probability is depending on the relative length of the edge w_i. This probability increases if the colour difference at the edge is larger than the mean, and conversely reduced if the colour difference is smaller than the mean.

Combining geometric and colour perception yields a quantity termed *composite perception*. The composite perception is denoted by P_C and given by

$$P_C = P_g\, P_c = P_g \frac{1}{\overline{\Delta E^*}} \sum w_i\, \Delta E_i^*\ , \qquad (2.51)$$

In Eq. (2.51) P_g denotes the degree of perception of the object in question due to its geometric relation with the observer; P_c denotes the perception due to colour difference between the object and its background. We note that since both quantities P_g and P_c are given by a probability the resulting quantity P_C is also given by a probability. That is, the model remains entirely as a probabilistic model. It is noted that the combination of the colour and geometric components is realized by means of multiplication of the probabilities. This is because both aspects are to be seen as independent probabilistic events.

2.5.3. Verification of the model by means of an experiment

In an experiment the composite perception computation described in the previous section is verified by means of *hypothesis testing*. The test persons were students of architecture. They were presented with two images shown in figure 2.27a and 2.27b. The computed colour differences at the borders of the patches are indicated in the figures. As display for the experiment a 24" high resolution computer monitor, brand HP, type LP 2465 was used. The distance between eye and screen was 1m.

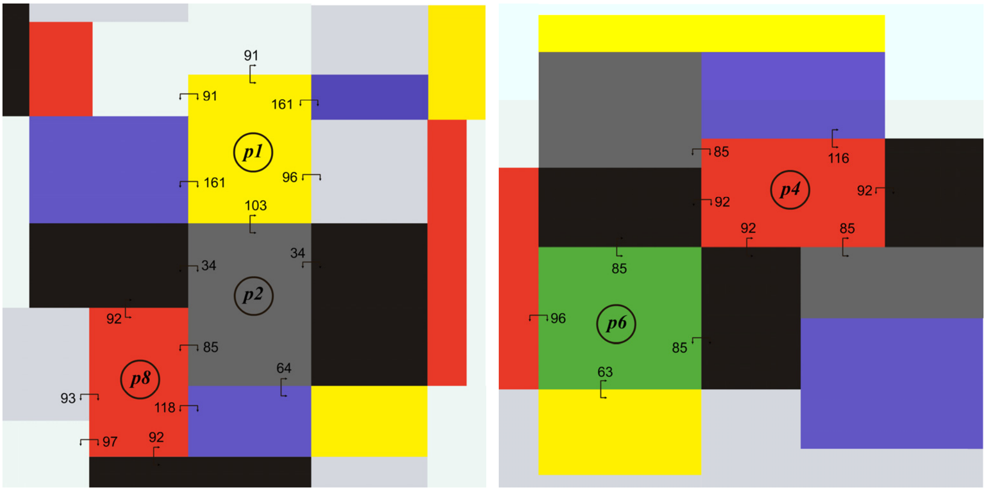


Fig. 2.27 Image used in the first perception experiment (a); in the second experiment (b); the colour differences at the borderlines of the patches that are subject to investigation are indicated.

This was done, so that the pdf of the geometric perception computation approaches to uniformity for computational convenience. The persons were asked to visually experience the image and identify the patch in the image that attracts their attention most. This was done once per person. For the first experiment 36 persons provided this data. The resulting frequency a patch was selected in the experiment is given in table 2.1 in the column labelled f_e. In the column labelled P'_C the composite perception is given. This is the probability a patch is selected as predicted by the model. It is noted that in the investigation the probabilities associated with patches other than the ones labelled p1 p2 and p8 are not considered.

The prime sign of P'_C indicates that the number is given normalized with respect to the scene in question. Normalization ensures that the two-dimensional probabilistic space consisting of a colour and a geometric dimension is narrowed down in accordance with the conditions of the experiment. These conditions are that an observer only selects from the presented combinations of colours and patches exclusively, while a large number of possible combinations are not subject to selection, although they formally are part of the probabilistic space $P_g x P_c$. The normalized composite perception P'_C given by a probability is converted to a frequency for comparison with the experimental frequency. This is done by multiplying P with the total amount of statements the persons made. The resulting theoretical frequency f_t is then compared with the experimental frequency f_e using the $\chi 2$ test method. The $\chi 2$ calculation is given by

Table 2.1 Statistical result of the perception experiment on the image in fig. 2.27a

patch	P'_C	f_t	f_e	$\frac{(f_t - f_e)^2}{f_e}$
p1	0.45	16.0	23	2.10
p2	0.28	10.1	7	1.36
p8	0.27	9.9	6	2.50
χ^2				5.96

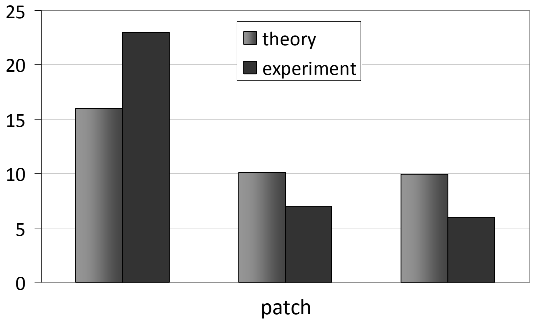


Fig. 2.28 Comparison of the theoretically predicted selection frequency and the experimental result for fig. 2.27a; the column pairs belong respectively to patch 1, 2, and 8, from left to right

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_t - f_e)^2}{f_e} \tag{2.52}$$

where *k* is the sample size. In the first experiment *k=3* for the 3 patches concerned. In figure 2.28 the theoretically predicted selection frequency is compared with the experimental results. The calculation of the *χ2* value is shown in table 2.1. The degrees of freedom in this experiment is *k-1=2*. The resulting *χ2* value is 5.96. This is slightly less than the threshold belonging to the critical region of size 0.05 of the *χ2* distribution for two degrees of freedom, which is 5.99. This confirms the impression from figure 2.28 that the model is acceptable.

In the same way a second experiment was conducted for the image shown in figure 2.27b. In the second experiment 43 persons provided the data. The theoretically predicted selection frequency is compared with the experimental results in figure 2.29. The result from hypothesis testing is presented in table 2.2. In the second experiment *k=2* for the two patches selected by the test persons. The degrees of freedom of the experiment are *k-1*, i.e. 1 in this case. The resulting *χ2* value is 0.49. This value is significantly less than the threshold belonging to the critical region 0.05 of the *χ2* distribution, which is 3.84. This means the second experiment strongly indicates the validity of the colour perception model proposed [66].

The results indicate the deviation between model prediction and experimental results are statistically insignificant. This means the model is acceptable. The results are interesting considering that the perception is a very soft issue.

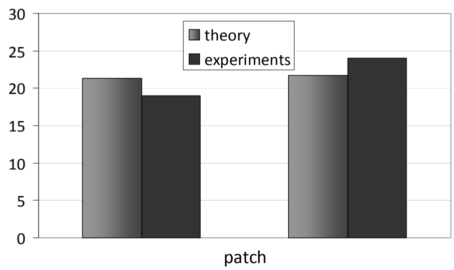


Fig. 2.29 Comparison of the theoretically predicted selection frequency and the experimental result for fig. 2.27b; the column pairs belong respectively to patch 4, and 6, from left to right

This means the statements of individuals may differ strongly among persons and even may change for the same person, when the person is asked at different moments. The individual preference of a person may bias the selection of a colour in the experiment and this bias may remain hard to identify. Despite this issue the perception computations apparently model the phenomenon. Also it should be emphasized that the experiments manifest the necessity to model the perception phenomenon probabilistically.

Table 2.2 Statistical result of the perception experiment on the image in fig. 2.27b

patch	P'_C	f_t	f_e	$\frac{(f_t - f_e)^2}{f_e}$
p4	.49	21.3	19	.27
p6	.51	21.7	24	.22
χ^2				.49

2.6 Applications of the colour perception model

2.6.1. Analysis of a retail interior

The colour perception computation above is applied for the perceptual assessment of an architectural scene. In the present case a retail environment is investigated. Specifically the perception of the entrance hall from its entry is investigated. The scene is shown in figure 2.30. In the scene a number of objects are subject to perceptual analysis. In particular the respective degree of perception per object is pursued. For this purpose the colour and geometric perception computations described above are used.

Fig. 2.30 Scene subject to composite perception analysis

Fig. 2.31 Segmentation of the scene shown in figure 2.30 for colour perception computation

For the colour difference calculation it is necessary to compute the circumference of the objects. This is seen from Eqs. (2.49) and (2.50). Obtaining the circumference may be automatically accomplished using image processing algorithms. In such algorithms edge detection is executed by identifying the locations of *zero-crossings* in an image [31]. Zero-crossings are the places in an image, where the second derivative of the function describing the level of brightness intensity for each location on the image plane, passes through zero. In other words, it marks the places where there is a peak in the function describing the intensity change. Detection of such places can be used to separate an image into segments. In order to identify zero crossings an image is processed by means of filters, which can be modified to operate on different length scales on the image to detect blur edges as well as sharp edges. Filters are commonly using a two-dimensional Gaussian function to blur the image together with the Laplacian operator to detect the zero crossings [67]. For our purpose it is sufficiently accurate to approximate the edges of objects by ellipses in fig. 2.31.

The calculation of the circumferences of the objects and the colour perception P_c at each object border is given in section A.5 of Appendix A. The different borderline segments are indicated in figure 2.31. The colour perception per object is obtained as the sum of the perceptions at each border of the respective object. The results are given in figure 2.32a. From the figure it is seen that from the colour perception viewpoint object O1 has the highest degree of perception. Objects O2 and O3 and O8 are also high. This can be verified by observing figure 2.29, where the yellow shopping bags stand out clearly.

In order to obtain the composite perception degrees of the objects the results from the colour perception computation are to be combined with the perception from geometric viewpoint. This is accomplished using Eq. (2.51). For this purpose the geometry based perception P_g of the objects in the scene is obtained. The result is presented in figure 2.32b. Comparing figure 2.32a and 2.32b we

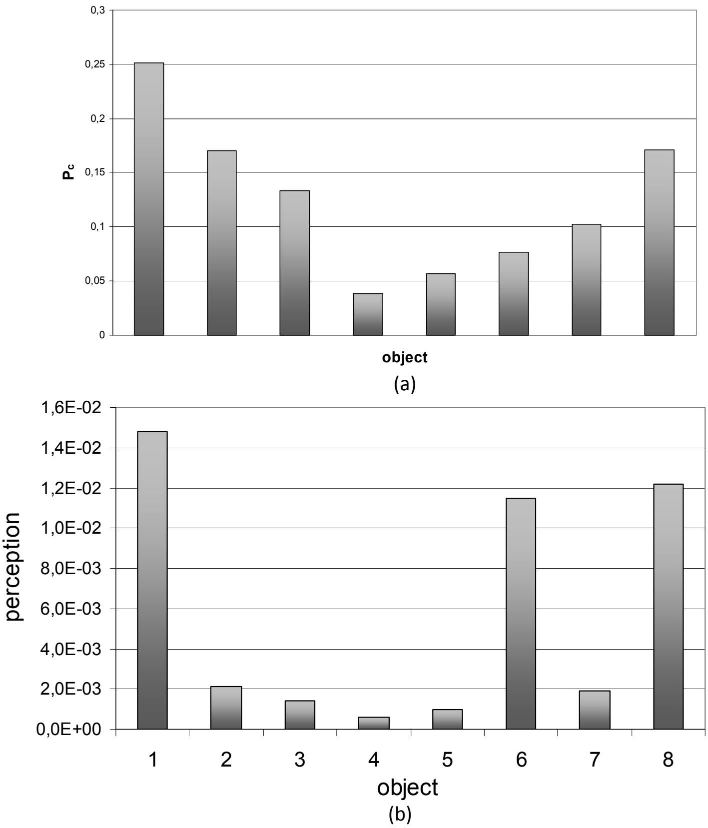


Fig. 2.32 Colour perception component (a); and geometric perception component (b) for the scene shown in figure 2.30

note that from the colour viewpoint the objects yield relatively similar perception, whereas in the geometric case the difference among objects is relatively stronger. The geometric perception measurements are conducted in a virtual reality environment. This is shown in figure 2.33, where a virtual observer is seen perceiving the objects of the scene.

The composite perception is given as normalized values in figure 2.34. One notes that the colour perception component in figure 2.32a is relatively less distinct among objects compared to the geometric component in figure 2.32b. From figure 2.32b it is seen that from a perception viewpoint the major objects of the scene are objects 1, 6 and 8. The perceptions of the other objects are quite marginal. When the perceptions of objects 2 and 3 are added up, i.e. when they are considered as one object, a notable perception value of .073 is reached.

It is interesting to compare figure 2.32 with the one in figure 2.34. From figure 2.32 we see that objects *O1*, *O6*, and *O8* have the highest perception from the geometric viewpoint. These objects clearly stand out with respect to the others. Considering figure 2.34 we note that this dominance prevails with respect to the composite perception P_C. Here object *O1* dominates the perception of the scene even more strongly than considering the geometric viewpoint only. Clearly this is due to the relatively higher colour perception of object *O1* compared to objects *O6* and *O8*. Especially the escalators' significant geometric perception is diminished due to the low colour difference between this object and its background.

Fig. 2.33 Implementation of the perception measurement in virtual reality

2.6.2. Perception along different viewpoints

From an architectural viewpoint it is relevant to gain insight into the experience of an observer that walks through the space. When will he note which object, and to what extent? To answer these questions the perception analysis is conducted along a trajectory in the retail space. For this purpose a commonly taken path of visitors is selected. The path is indicated as a blue line on the floor in figure 2.35. The virtual observer is programmed to follow the path with a constant speed. At distance intervals that correspond to 30 cm in reality, the unbiased perception degree is computed for each of the eight objects. The results are shown in figure 2.36. Distances in figure 2.36 are given as the number of 30 cm intervals.

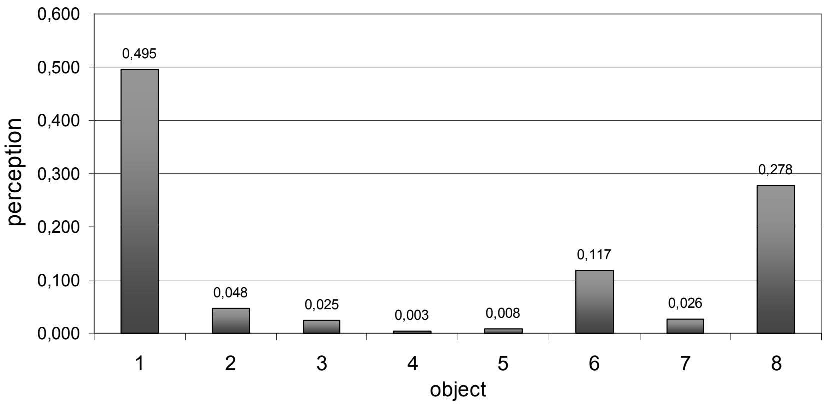


Fig. 2.34 Composite perception of the objects in figure 2.30

Fig. 2.35 Perceptual analysis along a trajectory in the space

From figure 2.36 it is seen that during first part of the trajectory, from interval 1 until 20, the awareness of the unbiased observer is 'dominated' by object *O1*. This object contains the yellow shopping bags.

During the second part, from interval 20 until 36, the observer becomes most aware of object *O8*, which are the exhibited furniture pieces. This second part is less strongly dominated, so that relatively more attention is paid to the other objects compared to the beginning of the trajectory. Clearly the final part of the trajectory accentuates the escalators (object *O6)*, more and more. In the meanwhile *O7*, i.e. the advertisement signage indicating prices at the restaurant of the store, is also receiving significant attention. In the intervals 42-47 the unbiased observer is even slightly more aware of the signage than of the escalators. Relocation of the objects in the scene will change this perceptual pattern

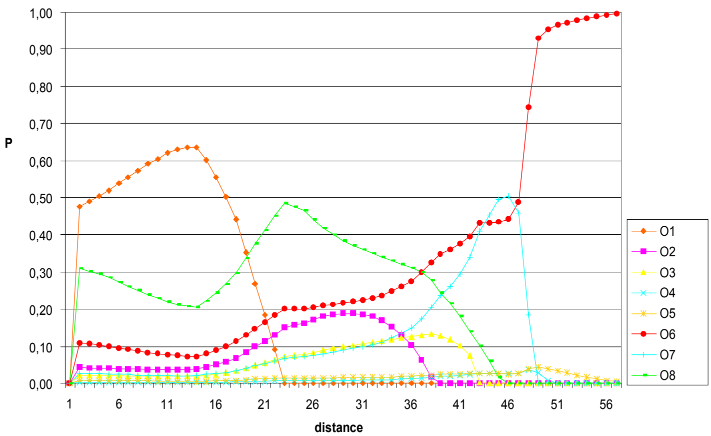


Fig. 2.36 Resulting perceptions along the trajectory for the respective objects

along the same trajectory. An objective in architectural design is to provide a perceptual pattern, i.e. a sequence of experiences that suits an intended purpose of the building. The perception model is useful to verify a pattern during a design process.

2.6.3. Analysis of an urban situation

In the same way as in the previous example a perceptual analysis of an urban square is conducted. In this case the degree of perception of a number of large scale advertisement displays is compared. The scene is shown in figure 2.37. Figure 2.38 shows the implementation of the measurement. The resulting perception components of the objects viewed by the virtual observer are shown in figure 2.39, where the constituents of the perceptions, namely the colour and geometric perceptions, are given respectively in figures 2.39a and 4.39b. The resulting composite perception is shown in figure 2.40 as normalized values.

From figure 2.40 it is seen that object *O1* dominates the perception with respect to the other objects. Comparing figure 2.40 with figure 2.39 it is seen that the resulting perception, where colour and geometric perception are combined, appears proportional to the geometric component. At the same time the composite perception accentuates the relative differences among the geometric perceptions of the objects even more. This is because in the present scene the background for the objects is similar for each object. Therefore the weight factor w in Eq. (2.50) determines the relative difference among the objects as to colour perception. Since in the present case objects with a relatively long circumference are also objects that have a high perception from the geometric viewpoint the combination of both factors increases perception strongly. It is emphasized that the influence of the shape proportions of an object stems from the definition of w_i in Eq. (2.50). In Eq. (2.50).the length of an object's borderline is normalized based on the sum of all borderlines present in the scene. This means an edge that is relatively long compared to the average edge-length has a relatively high w_i value. Therefore, in the present case the composite perception accentuates the differences from the geometric viewpoint.

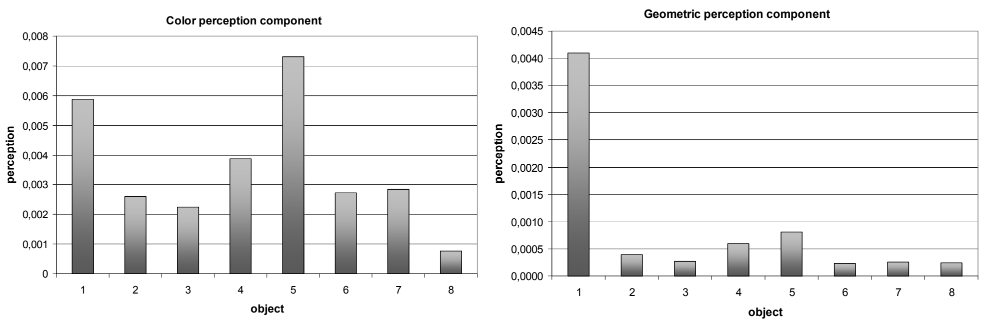
2.6.4. Perception from an alternative viewpoint

Let us investigate how the perception of the objects changes when the virtual observer assumes another viewpoint that is located beyond object *O1*. The second viewpoint is indicated as *P2* in figure 2.41. We note that due to normalization the resulting perceptions for the visible objects increase compared to the previous case. This is because a large portion of visual attention is no longer paid to object *O1.*

From figure 2.42 it is seen that object *O5* now dominates an unbiased observer's perception, while object 4 and 7 also gained significantly. In particular object *O7* gained relatively more perception than object *O4*, which is due to the closer position of *O7* to eye height and its greater proximity to the observer compared to *O4*.

Concerning object *O8* it should be noted that when the virtual observer is positioned closer to the object the perception increase for *O8* is not significant compared to the other objects. This is despite the fact that *O8* has some perception viewing it from a distance. This is because from a nearby distance the angles between the central axis of the virtual human's visual scope and the sightlines to the object become very large. Therefore attention diminishes strongly. From the analysis it can be concluded that the advertisement shown on the display *O8* is rather ineffective. This may be unexpected as we may intuitively think that an advertisement 'overlooking' the entire square should have a high perception.

Fig. 2.37 A number of objects in an urban scene subject to analysis regarding their unbiased perception



(a) (b)

Fig. 2.38 Colour and geometric perception components for the perception situation in figure 2.37

Fig. 2.39 Measurement of unbiased perception from position *P1*

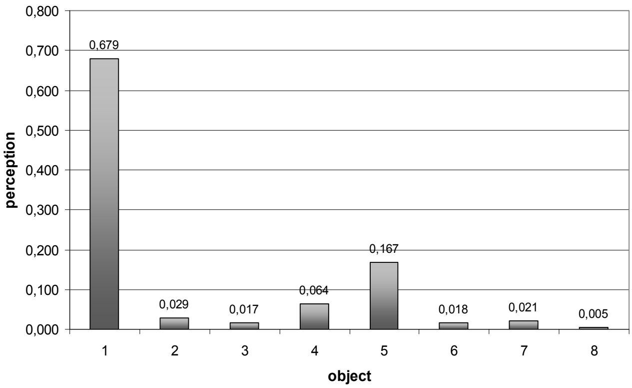


Fig. 2.40 Composite perception of the objects indicated in figure 2.37

Fig. 2.41 Measurement of unbiased perception from position *P2*

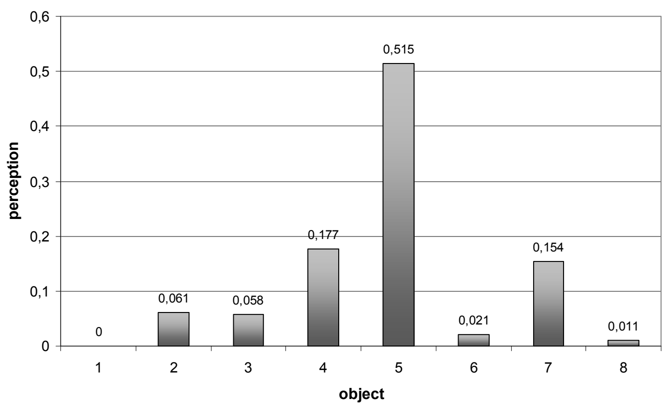


Fig. 2.42 Composite perception from position P2

2.7 Discussion

The perception theory presented is used for perceptual assessment of designs. This means designers can compare the perception properties of spaces with great precision. Using this information we are able to modify a design in such a way that it has the properties we aim for. With this goal in mind it is relevant that the validity of the model is not restricted to specific environments. This means the model should work as a standard perception evaluation metric for various environments.

Although an observer's visual perception may be influenced by contingencies, such as personal preferences or task induced bias, the results from the unbiased model are a significant source of information for a designer. Based on the results from the model a designer is able to put forward requirements for degrees of perception he deems desirable. In this way his own judgement of the circumstances is taken into account, while making at the same time use of the objective computations. The role of the computation is to reduce the complexity of the visual inspection task, where many factors usually bias an observer, so that his perception assessment is generally imprecise. The imprecision is reduced by using the model to acquire the perception information of the scene.

Next to geometry and colour other circumstances may influence vision probability. Examples are when an observer is searching for an object, or when he/she has a preference for the object for other reasons. Such a bias may increase or decrease the unbiased vision probability. In perception experiments, to identify such bias without information on basic influence of geometry or colour in perception is formidably difficult. That is, it is not to be traced which component of a vision probability is due to the basic influence, and which component stems from other factors. In this sense the present perception model for unbiased perception reduces the complexity of the modelling task, so that detailed insight into the influence of higher-level aspects may be obtained if this is desired.

The validity of the theory is verified by means of two vision experiments. Further manifestations of the probabilistic theory are found in the literature, where another common vision phenomenon is uniquely explained based on the probabilistic perception theory [6]. This phenomenon is the blur in human vision.

The validity of the theory presented is also confirmed as the results coincide with several manifestations we commonly experience: An observer is usually more aware of objects that are (i) located in the frontal region of our visual scope vs. side regions, (ii) that are nearby vs. distant, or (iii) that are large vs. small (iv) that have a higher colour contrast with their background. It is a basic common experience that these properties increase perception of an object. In this respect it is important to mention firstly that the corroboration of the theoretical results with daily experience is a significant indication of the validity of the theoretical considerations. Secondly, the results provide some insight into the mechanism responsible for the common phenomenon. And thirdly the relevance of distance on perception is quantified, as opposed to a vague verbal statement, e.g. that "perception decreases as distance increases."

We intuitively know that the closer an object is located from us the more aware we are of it. However, in contrast to what one may suspect in a general case reducing the distance to an object to half does not increase perception by factor 2 per space dimension. That is, our awareness of the object does not become twice as high. This means that the amount of visual detail obtained from an object does not increase proportional to the viewing distance in general. The relation is non-linear in general and subject to computation using the theoretical considerations.

It is important to note that the numeric result from the computation of unbiased vision probability is equivalent to the solid angle of an object in the visual scope normalized by the angle of the scope. Already J.J. Gibson conceived his *ambient optic array* as a set of solid angles [20]. However, of course, the calculation of a solid angle that is normalized by the angle of the visual scope by itself is an ad-hoc proposition as a model of perception. For example it is difficult to justify a-priori why distance between object and observer or the orientation of an object would not matter. But more importantly it is unclear what the resulting number actually means. So the major contribution of the model presented in this work is that it defines perception as a *probability*; namely the probability that the corresponding environmental information is mentally realized. The validity to take the normalized solid angle as measure of perception is confirmed by the theoretical considerations behind the present work, which is based on its premise and further derivations.

There are two alternative ways to calculate perception. One is based on the integral of the attention, which is a pdf, or based on the solid angle computation mentioned above. These calculation principles bear some difference. One difference is that the solid angle approach merely tells about the vision probability of an object in the form of a scalar number. In contrast to this the probabilistic approach involves the concept of *attention*, which is a probability density that is given per unit length and in a certain direction. Therefore, the orientation of the surface of an object is neglected in the solid angle computation as well as its distance. In the probabilistic case this information is expressed by a probability density in space. Certainly the probability density function, i.e. attention, may be approximated by subdividing the object into smaller patches and calculating their respective normalized solid angles. However, this may require ample computation depending on the accuracy needed in discrete form.

This means the solid angle model is useful to compute the vision probability of an object, if one is not interested how the perception density is varying throughout the visual scope or within the boundaries of an object. Especially the directions in the view of an observer, which are more prominent in perception sense, are given by the pdf and not by the probability. In architectural design the variation of perception density is valuable information, because in the design of a space it may be relevant which locations in the space receive more or alternatively less attention, or which directions are most prominent in this respect.

It is important to note that in the perception computations depth information is considered. Namely the depth is denoted by means of the parameter l_o in the equations for surfaces of objects that are oriented perpendicular to the forward looking direction. For objects extending in directions other than the perpendicular one, the pdf of perception is treated as a general case in this publication [15].

Concerning other aspects influencing perception, next to colour difference and geometry, a topic frequently articulated in the perception domain is *texture* of an object. As to the role of texture in perception firstly it is necessary to establish what is meant by texture. Originally texture refers to specific geometric variation on an objects surface yielding a corresponding experience of the surface when it is touched, that gives more information about the objects material. For example textures can be termed *rough, smooth, grainy* or *coarse, for* example. These variations can also be experienced visually, where the variations on the surface geometry imply a certain interaction with light reflected on the surface, which specifies the objects material [20]. Essentially the geometric variations and are variation in the atomic structure of the material, so that the atomic structure of the objects surface, i.e. its surface material, results in a specific pattern of light energy reflected per differential angle in visual perception. Having said this it is noteworthy to mention that the properties of the light source shining light onto a surface of concern are important aspects influencing the visual perception of the surface, too, next to the texture of the object. It is emphasized that texture is a concept used to specify a material. It should therefore be regarded as a sub-domain or modality of a material being a physical entity. Without a material we cannot talk about the texture of it, while we may have a material and we may not consider its texture. It is further clear that a material is what yields the experience of colour, as the experience of colour is due to a flux of energy having specific properties with respect to its spectral power distribution, which stems from the specific reflectance/absorbance behaviour of the object's surface material to light impinging on it. Therefore texture should be regarded as a sub-domain or modality of colour. This means, saying two objects have *different textures* implies that they have *different colours*. The difference is the modality of the colours. They can both be black for example, but the blackness maybe different due to material difference with its texture difference. We can talk of a *mat black* or a *glossy black* colour for instance. Such detailed effects of colour modalities on visual perception are an interesting relevance for further research.

Existing models of attention are not based on probability theory. They are statistical or deterministic in nature with their inherent shortcomings mentioned above. Therefore existing models should be brought to the probabilistic domain for integration. Concerning possible extensions of the probabilistic approach to perception presented, candidate features to be included into the perception model are features present in a scene that influence the probability to see objects. Such features are modalities of colour, such as texture difference, or pattern difference, and also modalities of geometry, such as shape difference, as well as task specific preference of an observer. Criteria to include such features are if these features are significantly different for the objects on the scene, and if they can be modelled in probabilistic terms. Also the desired accuracy of the model outcome is to be considered. It is noted that computation time may increase drastically as the amount of features subject to simultaneous consideration increase.

One may wonder if in the application examples presented *luminance*, being an important aspect in perception is considered. Luminance is the luminous intensity per unit area. It is a photometric quantity generally associated with light sources. When considering the light reflected from a surface we consider the photometric quantity termed *luminous excitance* being the luminous flux per unit surface area [68]. Both photometric quantities are considered in the present model by means of a single psychophysical quantity, namely the luminous intensity *Y* that is used to specify a colour in the C.I.E. 1931 *XYZ* colour space. It is noted that light, whether arriving at the eye from a source or via an object, can be specified by the CIE *x,y* coordinates together with a value describing the light's luminous intensity [69], where the *y* coordinate is termed luminance. This explicit expression of luminous intensity is accomplished by selecting one of the matching functions to be the luminous efficiency function of the standard photopic observer. This way C.I.E. arranged that luminance is explicitly given by the *y* coordinate. This is further described in the Appendix. For instance comparing light beams that have equal luminous intensity Y, green or yellow light is perceived as brighter than blue or red light, since they have a higher luminance, i.e. *y* value. It is noted that in the applications, instead of using XYZ specification CIE $L^*a^*b^*$ is used is used due to perceptual uniformity of the latter space, to obtain perceptually uniform colour difference. For convenience the colours obtained in the scenes in applications are obtained from digital photograph and not by means of colour measurement in the physical environment. As photographic image reproduction involves the matching functions of a human observer, that include the luminous efficiency function for the C.I.E. standard observer, luminance is included in the analysis of the images.

2.8 Conclusions

A novel theory of unbiased human vision is presented. In the theory human vision and related concepts are expressed using mathematical terms. Based on the theory perceptual properties of spaces are subject to analysis by means of computation. This is a significant step, because conventionally such aspects are considered to be too soft to be subject to computation. Using the computations an architect is able to compare different designs concerning their perceptual properties with precision. This is particularly helpful when visually rich environments are subject to comparison, where a large amount of visual information makes consistent visual assessments difficult otherwise. Such small differences in perceptual awareness may distinguish designs regarding their architectural quality. Therefore a building by Peter Zumthor, e.g., has a higher level of perceptual sophistication than conventional buildings. Next to this, considering perceptual aspects with precision is particularly beneficial when perceptual features are weighed against other design features. This is the case e.g. when requirements for perception properties, such as a high visual openness, are conflicting with other aspects of design, such as cost, functionality etc. In such a case tolerances can be exploited carefully to arrive at suitable solutions matching the preferences accurately.

The approach is uniquely based on probability theory. This is in contrast to earlier models of human vision. These were either deterministic [31] or statistical [26]. In the present theory attention is defined as a probability density. The integration of this *attention* over a spatial domain is *perception* and quantified by a probability. The probabilistic approach is uniquely able to cope with the complexity of the human vision process, where the complexity is essentially due to the involvement of brain processes. The theory corroborates with our common vision experience and its validity is confirmed by means of two vision experiments.

The perception model is generic in the sense that is not restricted to particular environments. Also model does not contain any information regarding perceptual preferences of an observer. It forms a standard measurement device that can be used in connection with contingent preferences during design. That is, an architect is able to bring his/her personal perceptual preferences into play and combine this subjective information with the objective results from the model. This is treated in the following chapter, where a method is described to interpret the results from perceptual analysis as satisfaction of design requirements.

2.9 References

1. Bittermann, M.S., Ciftcioglu, O.: Real-time measurement of perceptual qualities in conceptual design. Proc. Int. Symp. Tools and Methods of Competitive Engineering TMCE 2006, Ljubljana, Slovenia (2006) 231-239
2. Bittermann, M.S., Ciftcioglu, Ö.: Validation of a visual perception model. Joint Int. Conf. on Construction Culture, Innovation, and Management (CCIM), Dubai, United Arabian Emirates (2006) 289-299

3. Bittermann, M.S., Ciftcioglu, Ö.: Visual perception model for architectural design. Journal of Design Research **7** (2008) 35-60
4. Bittermann, M.S., Sariyildiz, I.S., Ciftcioglu, Ö.: Visual space perception model identification by evolutionary search. Design 2006, Dubrovnik, Croatia (2006) 185-192.
5. Bittermann, M.S., Sariyildiz, I.S., Ciftcioglu, Ö.: Visual perception in design and robotics. Integrated Computer-Aided Engineering **14** (2007) 73-91
6. Bittermann, M.S., Sariyildiz, I.S., Ciftcioglu, Ö.: Blur in Human Vision and Increased Visual Realism in Virtual Environments. Lecture Notes on Computer Science. Springer Verlag (2007)
7. Ciftcioglu, O., Bittermann, M.S., Sariyildiz, I.S.: Application of a visual perception model in virtual reality (poster). ACM SIGGRAPH Symp. Applied Perception in Graphics and Visualization, APGV'2006, Boston, USA (2006) 143
8. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Towards computer-based perception by modeling visual perception: a probabilistic theory. 2006 IEEE Int. Conf. on Systems, Man, and Cybernetics, Taipei, Taiwan (2006) 5152-5159
9. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Studies on visual perception for perceptual robotics. ICINCO 2006 - 3rd Int. Conf. on Informatics in Control, Automation and Robotics, Setubal, Portugal (2006)
10. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Visual perception theory underlying perceptual navigation. Emerging Technologies, Robotics and Control Systems International Society for Advanced Research (2007) 139-153
11. Ciftcioglu, O., Bittermann, M.S., Sariyildiz, I.S.: Fusion of perceptions for perceptual robotics. Proc. NAFIPS'06, Montréal, Québec, Canada. (2006)
12. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Autonomous robotics by perception. SCIS & ISIS 2006, Joint 3rd Int. Conf. on Soft Computing and Intelligent Systems and 7th Int. Symp. on advanced Intelligent Systems, Tokyo, Japan (2006) 1963-1970
13. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Sensor data fusion in autonomous robotics. The 2nd Int. Conf. Innov. Comp., Inf. and Contr. - ICICIC 2007. IEEE, Kumamoto, Japan (2007)
14. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Multiresolutional fusion of perceptions applied to robot navigation. Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII) **11** (2007) 688-700
15. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Further studies on visual perception for perceptual robotics. Fourth Int. Conf. Informatics in Control, Automation and Robotics - ICINCO2007, Angers, France (2007) 468-744
16. Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, O.: Perception & Architecture. In: Bekkering, H., Hauptmann, D., Heijer, A.d., Klatte, J., Knaack, U., Manen, S.v. (eds.): The Architecture Annual 2005-2006 Delft University of Technology. 010 Publishers, Rotterdam (2007) 104-109
17. Sariyildiz, I.S., Stouffs, R., Tunçer, B.: ICCT influence on spatial planning, design and construction process. Information and Communication Technology (ICT) in the Practice of Building and Civil Engineering. Association of Finnish Civil Engineers RIL, Helsinki, Finland (2001) 137-142
18. Maslow, A.H.: Toward a Psychology of Being. John Wiley (1999)
19. Foster, J.: The Nature of Perception. Oxford University Press (2000)
20. Gibson, J.J.: The Ecological Approach to Visual Perception. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey (1986)
21. Merikle, P.M., Daneman, M.: Conscious v. unconscious perception. In: Gazzaniga, M. (ed.): The New Cognitive Neurosciences. The MIT Press, Cambridge, MA (2000)
22. Adelson, E.H., Bergen, J.R.: The Plenoptic function and the elements of early vision. In: Landy, M., Movshon, J.A. (eds.): Computational Models of Visual Processing. MIT Press, Cambridge (1991) 3-20
23. Bertero, M., Poggio, T.A., Torre, V.: Ill-posed problems in early vision. Proceedings of the IEEE **76** (1988) 869-889
24. Papathomas, T.V., Chubb, C., Gorea, A., Kowler, E.: Early Vision and Beyond. MIT Press (1995)
25. van Hemmen, L., Cowan, J., Domany, E.: Models of Neural Networks IV: Early Vision and Attention. Springer (2001)

26. Knill, D.C., Richards, W.: Perception as Bayesian Inference. Cambridege University Press, Cambridge (1996)
27. Robinson, H.: Perception. Routledge (2001)
28. Levine, M.W., Sheffner, J.M.: Fundamentals of Sensation and Perception. Addison-Wesley, London (1981)
29. Palmer, S.E.: Vision Science. MIT Press, Cambridge (1999)
30. Bigun, J.: Vision with Direction. Springer Verlag (2006)
31. Marr, D.: Vision. Freeman, San Francisco (1982)
32. Poggio, T.A., Torre, V., Koch, C.: Computational vision and regularization theory. Nature **317** (1985) 314-319
33. Gazzaniga, M.S.: The Congnitive Neurosciences III. MIT Press (2004)
34. Julesz, B.: Binocular depth perception without familiarity cues. Science **145** (1964) 356-362
35. Cataliotti, J., Gilchrist, A.: Local and global processes in surface lightness perception. Perception & Psychophysics **57** (1995) 125-135
36. Ghosh, K., Sarkar, S., Bhaumik, K.: A possible explanation of the low-level brightness-contrast illusions in the light of an extended classical receptive field model of retinal ganglion cells. Biological Cybernetics **94** (2006) 89-96
37. Desolneux, A., Moisan, L., Morel, J.M.: A grouping principle and four applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 508-513
38. Pentland, A.P.: A new sense of depth. IEEE Trans. on Pattern Analysis and Machine Intelligence **9** (1987) 523-531
39. Favaro, P., Soatto, S.: A geometric approach to shape from defocus. IEEE Trans. on Pattern Analysis and Machine Intelligence **27** (2005) 406-417
40. Subbarao, M., Choi, T.: Accurate recovery of three-dimensional shape from image focus. IEEE Trans. on Pattern Analysis and Machine Intelligence **17** (1995) 266-274
41. Turvey, M.T., Shaw, R.E.: Ecological foundations of cognition. Journal of Consciousness Studies **6** (1999) 95-110
42. Benedikt, M.L., Burnham, C.A.: Perceiving architectural space: from optic rays to Isovists. In: Warren, W.H., Shaw, R.E. (eds.): Persistence and Change. Lawrence Erlbaum Associates, London (1985)
43. Batty, M.: Exploring isovist fields: space and shape in architectural and urban morpohology. Environment and Planning B: Planning and Design **28** (2001) 123-150
44. Franz, G.M., von der Heyde, M., Bülthoff, H.: Predicting experiential qualities of architecture by its spatial properties. In: Martens, B., Keul, A. (eds.): Designing Social Innovation: Planning, Building, Evaluating. Hogrefe and Huber, Cambridge, MA (2005) pp. 157-166
45. De Floriani, L., Marzano, P., Puppo, E.: Line-of-sight communication on terrain models. Int. J. Geographical Information Systems **8** (1994) 329-342
46. Turner, A., Doxa, M., O'Sullivan, O., Penn, A.: From Isovists to visibility graphs: a methodology for the analysis of architectural space. Environment and Planning B: Planning and Design **28** (2001) 103-121
47. Herrero, P., Greenhalgh, C., Antonio, A.d.: Modelling the sensory abilities of intelligent virtual agents. Autonomous Agents and Multi-Agent Systems **11** (2005) 361-385
48. Hubel, D.H.: Exploration of the primary visual cortex 1955-78 (Nobel Lecture). Nature 299 (1982) 515-524
49. Wiesel, T.N.: Postnatal development of the visual cortex and the influence of environment (Nobel Lecture). Nature **299** (1982) 583-591
50. Cowey, A., Rolls, E.T.: Human cortical magnification factor and its relation to visual acuity. Experimental Brain Research **21** (1974) 447-454
51. Taylor, J.G.: Neural networks of the brain: Their analysis and relation to brain images. Int. Joint Conf. on Neural Networks IJCNN 2005, Montreal, Canada (2005)
52. Hecht-Nielsen, R.: The mechanism of thought. IEEE World Congress on Computational Intelligence WCCI 2006, Int. Joint Conf. on Neural Networks, Vancouver, Canada (2006) 1146-1153

53. Laberge, D.: Networks of Attention. In: Gazzaniga, M. (ed.): The New Cognitive Neurosciences. The MIT Press, Cambridge, MA (2000)
54. Rensink, R.A., O'Regan, J.K., Clark, J.J.: To see or not to see: The need for attention to perceive changes in scenes. Psychological Science **8** (1997) 368-373
55. O'Regan, J.K., Deubel, H., Clark, J.J., Rensink, R.A.: Picture changes during blinks: looking without seeing and seeing without looking. Visual Cognition **7** (2000) 191-211
56. Treisman, A.M.: How the deployment of attention determines what we see. Visual Cognition **14** (2006) 411-443
57. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology **12** (1980) 97-136
58. Ciftcioglu, O., Bittermann, M.S.: From perceptual towards cognitive robotics in the framework of evolutionary computation. In: Pennacchio, S. (ed.): Emerging Technologies, Robotics and Control Systems. International SAR, Palermo, Italy (2009)
59. Papoulis, A.: Probability, Random Variables and Stochastic Processes. McGraw-Hill, New York (1965)
60. Posner, M.I., Petersen, S.E.: The attention system of the human brain. Annual Review of Neuroscience **13** (1990) 25-42
61. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence **20** (1998) 1254-1259
62. Allport, A.: Visual Attention. In: Posner, M.I. (ed.): Foundations of Cognitive Science. MIT Press, Cambridge, Massachusetts (1989)
63. James, W.: The principles of Psychology. Thoemmes Continuum (1890)
64. Ciftcioglu, O.: Shaping the Perceptual Robot Vision and Multiresolutional Kalman Filtering Implementation. In: Pennacchio, S. (ed.): Emerging Technologies, Robotics and Control Systems InternationaSAR Publishing, Palermo, Italy (2008)
65. CIE: Colorimetry - Publication CIE 15.2-1986. Central Bureau of the Commission Internationale d'Éclairage, Vienna (1986)
66. Hoel, P.G.: Introduction to mathematical statistics. Wiley, New York (1984)
67. Marr, D., Hildreth, E.: Theory of edge detection. Proceedings of the Royal Society of London, Series B (1980) 187-217
68. Ohata, N., Robertson, A.R.: Colorimetry. Wiley, Chichester, UK (2005)
69. Nassau, K.: The Physics and Chemistry of Colour. Wiley, New York (2001)

3. Design performance evaluation

In the previous chapter a model of visual perception was presented. Requirements for perceptual properties, like most requirements in design, are generally not sharply defined. This means a designer tolerates this to some extend when such requirements are not fully satisfied. The tolerance is because requirements are conflicting. Therefore, when some satisfaction is sacrificed on one aspect it is expected that some performance is gained regarding another aspect as trade-off. Requirements that are not sharply defined need special computational treatment. This is a major issue treated in the present chapter.

In this chapter a novel method to model design requirements is described that is used to evaluate the performance of alternative designs. In particular the method provides the overall performance score of a design with the presence of multiple requirements that are not sharply defined. This chapter is based on the publications [1-7].

In section 3.1 the concept of performance is explained and the challenging issues to accomplish performance evaluation of architectural designs are described. In section 3.2 existing methods for performance evaluation are discussed, identifying the drawbacks for their use in architectural design. In section 3.3 a novel method for performance evaluation is presented. It is able to model the complex relations between design variables and design performance. The method's favourable properties for use in architectural design are pin-pointed.

3.1. The concept of performance

3.1.1 Introduction

When we are designing we frequently observe our design asking ourselves, or other experts, how good it is. This means we assess how well the design will *perform* once it is built. In order to understand how good the overall performance of our design is we look at several aspects of it. For example we examine its performance regarding the aspect of *functionality*. This means we check if the spaces have the right size and proportions for their intended use, if the functions that belong together are in proximity to each other, and if the flow of people in the building is efficient. We also consider its technical performance. This means we verify if the load-bearing structure of the building resists effectively and efficiently to the expected live loads, dead loads, wind loads, and earthquake induced forces. We verify if the building will lose little heat energy over time, and if the solar gains are exploited extensively; if the installation system effectively and efficiently provides the desired indoor climate; and if the amount of daylight is suitable for the intended use scenarios. We also verify if the expected construction and maintenance costs are low enough. Recently another technical aspect becomes relevant, which is the consideration if the building components can be reused or recycled to a great extent.

The functional and technical performance aspects address the primary human needs for safety and shelter. Next to these, higher-level human needs also need to be considered. This means we verify if the building will satisfy the clients' and users' esteem needs and needs for self-actualization. For example we verify if the shape and materialization of a building expresses a client's personal or corporate identity. We also check if its spaces possess the perceptual properties suiting the user's preferences. This means we observe e.g. if the degree of visual openness and visual privacy are adequate, if the materials, the detailing and furniture of the interior provide the suitable ambience, and if the views to the exterior enhance the experience of the interior. Essentially we aim to predict, as good as we can, if a user or client is able to develop themselves in their preferred direction through the environment we provide. These high ambitions regarding the satisfaction of esteem and actualization needs of an individual are complemented by the aim to satisfy the esteem needs and needs for self actualization of our societies. This means a building is not only expressing the personality of a client or future user, it also expresses the state of our culture. This is clear considering that an individual exists in a context of people, and so does a building exist in a context with other buildings and urban elements. Beyond the immediate built context surrounding a building there is another, immaterial context in which we build and to which our building will react and contribute. This is the Zeitgeist, which reflects cultural, economic, and technological development. At certain times the zeitgeist 'prefers' buildings that express clarity and soberness, as it was the case during renaissance and modernism. At other times buildings are demanded that express vividness and convolutions such as during Baroque or in the expressionism. A building reacts and acts to the preferences of the zeitgeist in which it is being built. An architect is expected to be aware of such aspects and take them into account considering the appropriateness of his/her design, i.e. during performance evaluation.

These various performance aspects, functional performance, technical performance, energy performance, aesthetic performance, cultural performance, etc. together make a value we can term as *design performance*. This means we can define design performance as the total degree of satisfaction of the design requirements, which reflect a design's intended purpose. Requirements may include desirable features put forward by private stakeholders, such as clients and users, as well as imperative features demanded by law. One notes that that the term requirement is used in this thesis in a broad sense, so that it encompasses the concepts of desire and demand, which may differ in associated degrees of tolerance for their satisfaction.

The performance concept is frequently articulated in building design [8-11] and its relevance is recognized [12]. Despite that it is generally loosely defined, and it is not made explicit, how a design's performance is evaluated precisely. There are many tools available to assess use of energy, thermal flow, structural behaviour, life cycle costs, etc. [13]. Each of these aspects influences some performance aspects of a design. However it remains unclear to what extent each of them influences the *total* performance. That is, it is not obvious how to combine the results from assessments of performance aspects to obtain information on the *overall performance*. For example it is problematic to tell how suitable a design is when it is using little energy, but it does not function well, or what the performance of a design is that has a low cultural value while being highly functional. The overall performance is essential information in design, because it allows us to compare alternative designs and to arrive at designs that are suitable in a comprehensive sense. The difficulty to assess the performance of a design stems from the fact that performance aspects, such as technical performance and cultural performance, are complex concepts, i.e. they are linguistic in nature, making it is problematic to consider them in perspective and with precision.

Lack of adequate means for performance evaluation forms a major bottleneck when assurance of design quality is aimed for. That is, decision makers are left with rather high uncertainty concerning the extent a proposed design suits the intended purpose of their project. They are largely unaware of the trade-offs characterizing the design task, so that they have too few information to steer the design process towards designs that match their intended purpose adequately.

Also, performance evaluation forms a crucial bottleneck with regards to development of advanced computational design methods and ensuing design technology. This is because the search for solutions with suitable performance features requires a model to grade the performance of alternative solutions.

3.1.2 Complexity of design criteria

The essential task of performance evaluation is to provide the degree of overall design performance given the satisfaction of elemental requirements. To accomplish this task models need to be established that capture the relation between design variables and total performance. This is challenging due to the complexity of the performance criteria involved.

Due to this complexity it is not known precisely what influence a modification of a certain design variable has on the overall performance of a design. For example if we consider a simple design modification, say changing the location of the entrance, or choosing a different construction material for the columns of a building, it is merely vaguely known what the implications of such modifications are with respect to the total design performance. Is the design getting better or worse? And how much better or worse is it getting? When a design object is modified the modification usually affects the performance on multiple aspects. It is difficult for a decision maker to be aware of these implications for the overall design performance. This difficulty stems from the complexity of the criteria.

The first source of the complexity of criteria is the *high amount of attribute relations* involved in performance evaluation. For example a *good* design may need to be highly functional, sustainable, and inexpensive. But what is meant e.g. by being *functional* is subject to clarification from case to case. It may mean different things in different contexts and for different clients. Generally we can consider it is an attribute that consists of a number of sub-attributes, which are themselves complex. For example, the functionality of a building may comprise the provision of sufficient floor space for the required activities in the building, as well as the accessibility among the spaces. But what is meant by *sufficient* floor-space, and what is meant by *accessibility* is again subject to clarification. Next to this it is usually not made explicit how strongly functionality depends on other aspects, or to what extent functionality plays a role with respect to the design performance, compared to the other factors involved.

Due to the high amount of relations among performance aspects and their different kind it is challenging to establish a model where the aspects are integrated forming a coherent whole. The bottleneck apparently lies in the finiteness of the human ability to resolve detail and store information [14]. Generally speaking it is difficult for an architect to concentrate on detailed issues of a design, while maintaining an overview of the entire design at the same time with its ample solution options. This dilemma applies to individual persons as well as a group of cooperating persons, where the different individuals govern different performance aspects. Different experts in a design team have knowledge on the relations among sub-aspects of performance. For performance evaluation it is relevant to establish a model that absorbs pieces of performance-related knowledge from different individuals into a unified model. It is a formidable task for an individual without appropriate computational means to grasp the high amount of attribute relations inherent in the expert knowledge.

A second source of complexity of requirements stems from the fact that in design usually some requirements are conflicting. For example a building should be inexpensive, yet provide high comfort at the same time. Conflicting requirements means fulfilment of one requirement reduces fulfilment of at least one other requirement. Therefore conflicting requirements cannot be simultaneously fulfilled totally, i.e. their satisfaction is partial. We can speak of a *degree* of requirement satisfaction. For example we may wish to have a comfortable building that is also inexpensive. Since these criteria are conflicting to some extend clearly we will get a building with some degree of comfort and some degree of inexpensiveness, because we cannot have both criteria fulfilled totally.

Due to partial satisfaction, a design requirement is generally not sharply defined, but it involves some tolerance for imprecision. This makes the concept of requirement a *fuzzy* concept. The tolerance refers to the acceptance of some deviation of a design variable from its ideal value. For example a room may be required to have a floor area of ideally $20m^2$, but it would also be acceptable if it had 17 m^2. However, in the latter case the degree of satisfaction is clearly somewhat less compared to the ideal case. This means that the truth the requirement is fulfilled is merely partial for $17m^2$, i.e. it is less than 100%. We note that truth values associated with satisfaction of a requirement have a scale with possibly infinite truth values ranging between zero and one.

With this understanding it is interesting to note that in design most requirements are fuzzy to some extend. This property is particularly obvious concerning perception-based requirements. For example statements such as “the elevators should be easily noticeable from the entrance,” or “the space should have a high visual openness” are clearly not sharply defined. This means such requirements may be not fulfilled, totally fulfilled or anything in between.

However this fuzziness is not only characteristic for perceptual requirements. It also belongs to requirements that are conventionally considered to be sharp, such as requirements on the costs of a building. As an example let us imagine we are buying a laptop computer, and we have in mind it should cost 1000 Euro. We consider laptops that cost exactly this amount, and we also consider laptops that are somewhat below this price and above. When we look at a cheaper one we will observe how much performance we will have to trade in for the reduction in price, and when we look at a more expensive one we consider how much performance we gain for the increased price. This means we have to somehow make explicit how tolerant we are with respect to the exceeding of the ideal price for more expensive products. In other words we need to know how our satisfaction changes as the price increases. This means that our requirement for a suitable price usually has some tolerance for partial satisfaction making it a fuzzy requirement.

Another example for a requirement that is conventionally considered to be crisp but that in effect has fuzzy properties is structural safety. It is clear that a building has to fulfil certain basic requirements regarding the structural behaviour under certain worst case load conditions. These calculations include safety factors that account for imprecision in the models and other uncertainties. Beyond that however, if a client demands this, it is possible to make the structure safer than is legally required [15]. The gain in safety certainly comes at a price, namely increase in building costs, and possibly reduction of aesthetic values due to thicker member size for instance.

From the examples we note that the fuzzy nature of requirements originates from the fact that we can hardly pin-point a certain value for a design variable as a single permitted value. That is because requirements are conflicting, so that we have to remain flexible to some extend in order to exploit the potentials of a solution maximally, i.e. in accordance with our specific intended purpose. As every project is different, i.e. every client, user group, and every economic, political, cultural circumstance may be unique, it is clear that tolerances associated with requirements of a design are usually unique, too. Due the conflicts among requirements we are generally interested in a compromise that takes into account the specific tolerances regarding the fulfilment of each requirement. Therefore clearly we need to have information about these tolerances during design and make use of this information, i.e. to exploit these tolerances maximally.

It is emphasized that most design requirements are soft. This is valid except regarding features demanded by building laws and regulations. The softness is because most requirements involve multiple attribute relations and tolerance for partial satisfaction. Therefore methods for performance evaluation are needed that deal with softness of information. This is a challenging issue when existing methods for performance evaluation are used.

3.2. Existing methods for performance evaluation

Performance evaluation requires capturing the relation between the variables defining the design and a corresponding total performance value. Such a model is supposed to duly take care of the complexity of the performance criteria. This means the relation among the design variables and the overall design performance is established, while the commonly high amount of relations among the performance aspects and the fuzziness of requirement satisfaction are adequately treated.

3.2.1 Aggregation-based approaches

There are a number of works reported in the literature, in which performance aspects are evaluated using a weighted summation of degrees of satisfaction [16, 17]. Experts are expected to express the relative importance among the elemental requirements, so that performance of design ***d*** is calculated as

$$\boldsymbol{d} = \sum_{i=1}^{i=n} s_i \, w_i \ , \qquad (3.1)$$

where s_i expresses the degree of satisfaction of the i-th requirement and w_i expresses the relative importance of the requirement with respect to the design performance; n denotes the total number of requirements involved.

Accurate ranking of the relative importance ***w*** among a large number of requirements is difficult for a decision maker. One reason for this are human memory limitations [14, 18, 19], provided that a decision maker is aiming to exercise comparisons among all elemental requirements. When a decision maker aims to put a certain requirement in perspective with the others it necessitates that he/she bears the other requirements in mind as well including their associated relevancies while making the judgement. Clearly this is a formidable task even for a moderate number of requirements. Therefore the weights are generally determined with significant imprecision, while the weighted summation approaches have no tolerance for the imprecision. Therefore it is likely that the search for solutions with high performance is misguided.

The accuracy of specified weights can be increased to some extent by using certain mathematical techniques. This way an expert is required to specify relative importance merely *pair-wise* [20], as opposed to being forced to bear multiple requirements in mind. Such a technique is the analytical hierarchy process (AHP) [21]. AHP is based on gathering pair-wise information about the relative importance of aspects in a matrix, which is called reciprocal judgement matrix. The principal eigenvector of the matrix is obtained and normalized, and it contains as its components the ranking of the aspects. A desirable feature of the method is that it indicates the degree of inconsistency in the judgments, and it tolerates some imprecision in this respect. With increasing number of aspects the imprecision of the ranking of the aspects increases also, and it may become unacceptably high. The imprecision is due to natural inconsistency in the pair-wise judgement, where the context of variables subject to comparison plays a crucial role. This issue becomes apparent when certain elemental requirements are compared that belong to different performance aspects. For example it is clearly difficult for an expert to determine accurately the relative importance between the types of insulation used in the walls compared to the position of the front door, regarding their influence on the total performance. This is because the contexts of these variables are very different. One can say such a task is like comparing apples with pears. Considering that requirements in design projects are diverse, it is problematic to ensure consistency in the judgements using AHP. For this reason it is appealing to employ the AHP in a hierarchical structure of aspects [22], where a number of requirements are aggregated forming a higher level attribute. Using AHP in this way allows specifying more accurately the relative importance of aspects and sub-aspects. However, it does not address an essential issue when forming knowledge models for performance evaluation, as follows.

Next to human memory limitations, there is a second reason for the difficulty of weighted summation as a viable approach for performance evaluation in design. This issue concerns that the requirements subject to comparison are soft in nature. For instance it is problematic to compare the importance between *high functionality* and *high sustainability* of a building. This is because both concepts are generally vaguely defined as they are *linguistic* in nature. That is, they are concepts and their meaning depends on their specific relations with a number of design variables, such as geometric and material variables. Generally this relation exists via other linguistic concepts, such as *high accessibility*, *low energy use, suitable sizes of spaces*, etc. Therefore naturally the concepts' definitions involve uncertainty and imprecision, and their relations with design variables are non-linear and challenging to model. In particular it is emphasized that a computational system that is to hold knowledge should contain *non-linearity* to be able to do that. Weighted summation-based approaches are linear, and therefore they are unsuitable to model knowledge for evaluating design performance.

Examples of attempts to deal with performance issues using linear functions can be found in the literature. For example [23] aims to compute the performance by considering the hierarchical relations among performance. In this approach aspects that constitute performance are considered, and for each aspect a relative importance is assigned, so that the individual relevancies sum up to one. A share of importance assigned to a particular aspect is then divided among its sub-aspects, and further subdivided among sub-sub-aspects until the relative importance of an elemental requirement is specified. The same approach forms the essence of commonly used techniques to calculate sustainability related scores, such as *Breeam* or *Leed*. The results in these cases, which are termed *score*, *utility number* or *Nutzwert*, are obtained using the relative importance of an elemental requirement in a weighted summation as given by equation 3.1. It is emphasized that the linear character of the computation does not treat imprecision of the specified relevancies and it does not hold knowledge. This implies that if an expert specifies a weight with only slight inaccuracy, then the difference of the resulting performance value may be significant,

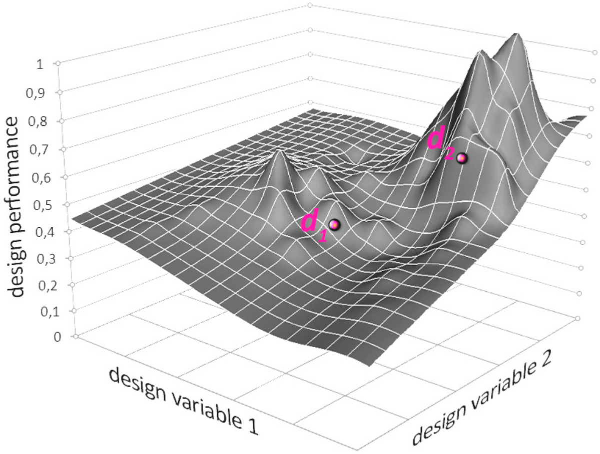


Fig. 3.1 Sketch showing a modelled knowledge for performance evaluation

depending on the circumstances. The size of the error regarding the resulting performance value depends on the particular set of input values used in the computation. In contrast to this, it is desirable that a model is formed, where the error inherent to the model is independent from the input values coming to the model. An error due to inaccuracy in the knowledge should be kept under control in the sense that it may not be drastic when a weight is slightly inaccurately specified.

As a summary of this section we note that the weighted summation-based approaches are unsuitable to model knowledge due to their linear nature and lack of tolerance for imprecision. This entails that their application is restricted to cases where the relations among aspects are linear, and known accurately and with certainty in advance.

3.2.2 Requirements for advanced approaches

Considering the shortcomings of the conventional computing approaches, clearly there is a need for more advanced treatment of the complexity of design criteria. In particular models are needed that are able to handle the partial satisfaction of requirements and that are tolerant for imprecise specification of the relations among performance aspects.

In the last decades a novel computing approach emerged that is based on exploiting the tolerance for imprecision inherent to complex real world problems. This way the approach achieves tractability and low solution cost. This approach is termed *soft computing* [24], which is a term coined by Professor Zadeh. Soft computing methodologies are also referred to as computational intelligence (CI) methodologies. These methodologies are the essential means to handle the imprecision that is inherent to human knowledge on a complex subject matter. This is because these methodologies tolerate imprecision in the information they process. This way an outcome obtained is not significantly affected by the imprecision. This means soft computing models are robust. Due to these properties soft computing methodologies are appealing to form models for performance evaluation in architectural design. Considering the soft computing methodologies, two major approaches for performance evaluation can be distinguished: knowledge-driven, and data-driven methodologies.

Before describing these approaches it is interesting to note that both of them produce a model of the relation between the variables of a design and a corresponding performance value. Such a model is referred to as a *knowledge model*, because it contains knowledge on how to measure the performance. Knowledge can be understood as means to arrive at a suitable result in a situation one has not previously encountered. The result is termed suitable in the sense that it is consistent with situations previously encountered, although they were different from current situation.

In mathematical terms a knowledge model corresponds to a *hypersurface,* that is a surface in a multidimensional space; namely it is an (*n*-1)-dimensional surface embedded in an *n*-dimensional space. The amount of dimensions *n* is the sum of the amount of input variables and output variables pertinent to the knowledge. That is the space in which the surface lies comprises the input variables, and output variable(s). The input variables are also referred to as independent variables and output variables as dependent variables. The surface is named as knowledge surface.

In case of performance evaluation the input variables are the design variables, and the output variable is merely one variable, namely the design performance. An example of a simple knowledge model is sketched in figure 3.1. The space consists of only two independent variables denoted as design variable 1 and design variable 2 in the figure.
The dependent variable is the design performance. In this example a certain design ***d*** has two values as inputs, and it yields a corresponding design performance value at the output. Two designs d1 and d2 are shown in figure 3.1, where the design d2 clearly has a higher performance than d1.

The knowledge surface ensures that for any input the model provides the unique corresponding location on the knowledge surface as an output value. That is, for a given set of values for the design variables, the model provides a corresponding performance value. It is noted that the relation among design variables and performance may be highly non-linear, and that the amount of independent variables of a design is generally much higher than two. It is noted that data-driven and knowledge driven methodologies are alternative ways to establish such a knowledge surface.

3.2.3 Data-driven approaches

Data-driven methodologies are based on 'distilling' knowledge out of data by means of *learning*. In case of use for performance evaluation the data are the values of the design variables together with a corresponding performance output value. These values have to be available for the model formation. The knowledge derived from a multitude of such data sets models the relation between design variables and performance.

The most prominent methodology for data-driven knowledge modelling is formed by artificial neural networks (ANNs). An ANN is a computational system that maps input vectors to output vectors. The mapping is accomplished by using processing elements termed *nodes* and connections between them, with weight coefficients associated to the connections. As an example the general description of a type of ANN that is known as *basis function network* is given by [25]

$$f(\boldsymbol{x}) = \sum_{i=1}^{n_\theta} \theta_i \, \phi_i(\boldsymbol{x}) \, , \tag{3.2}$$

where **x** is the input vector and f(**x**) is the output of the model. One notes that that ANNs may have multiple outputs as well. The network has n_θ nonlinear processing units termed

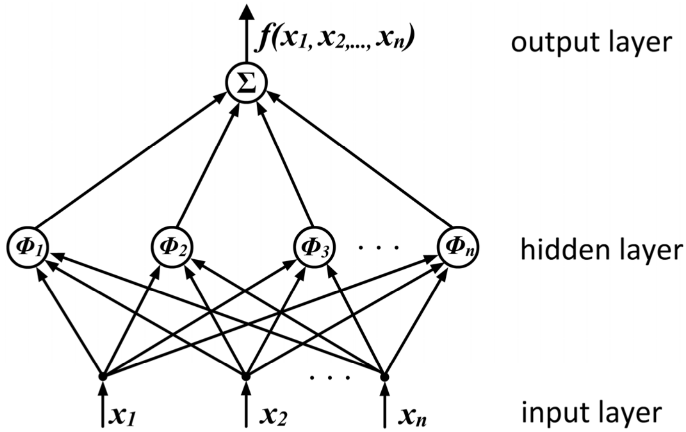


Fig. 3.2 Structure of a feed-forward artificial neural network

nodes. The nonlinearity of the *i*-th node is given by the function $\phi_i(\mathbf{x})$, where $\mathbf{x}$ is the input vector. The output of each node is multiplied by the weight θ_i and these values are summed up forming the output of the network.

The connection weights are the 'memory of the system.' ANNs are referred to as connectionist models due to the significance of the connections among nodes. The structure of the artificial neural networks is inspired by the constitution of the human brain as a network of neurons that are processing information in parallel. Each neuron of an ANN performs a non-linear mapping of an input vector to a scalar output. The non-linear functions involved often are selected as sigmoid or Gaussian functions. The output of one neuron is used as one of the inputs of other neurons. A common neural network type is the feed-forward network, where there are no connections back from the output to the input neurons. An example of a feed-forward ANN is shown in figure 3.2. The first layer is termed *input layer* and the last layer is termed *output layer.*

The intermediate layers are termed *hidden* layers, because the nodes on these layers generally have no associated comprehensible meaning. The neuronal computations together yield a complex computation.

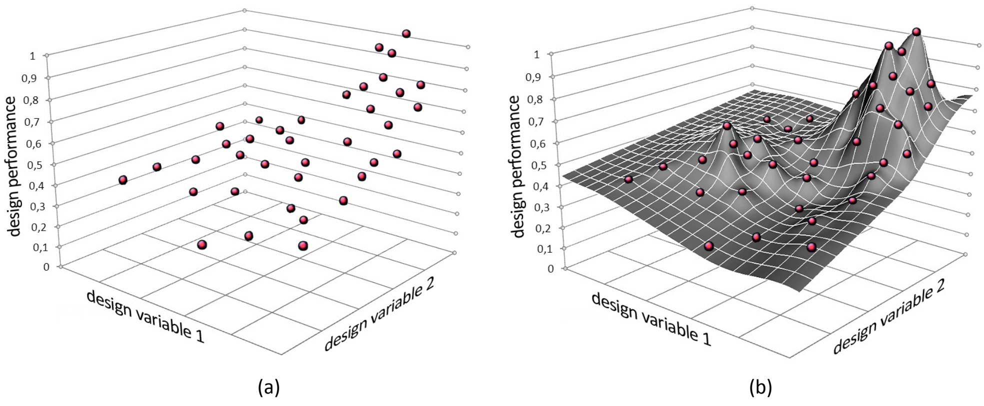


Fig. 3.3 Establishing a knowledge model: input-output data points (a); knowledge model obtained after learning (b)

Like the individual neuron, the network also performs a mapping of an input to an output, while the latter mapping is clearly more complex. This way the relation between input and output is permitted to be highly non-linear, depending on the complexity of the network structure and activation functions.

Certain types of ANN are universal function approximators. That is any non-linear function can be approximated with such networks. To prepare a neural network for performance evaluation of a design, the network is provided with a number of input data $\boldsymbol{d}$ and corresponding output data $\boldsymbol{p}$. This way the relation between them is modelled. In case of performance evaluation the input is a vector $\boldsymbol{d}$ that contains the values of the design variables. The output is a scalar number p that expresses the performance of the design.

The process of establishing the relation between input and output is referred to as *training* or *learning*. It is a process of establishing a general relation between an input and an output accommodating the relations inherent to the data sets provided. The learning is adjustment of the parameters of the network, so that the model represents the given set of input-output data pairs with minimal error. The parameters subject to identification concern the weights of the connections among the nodes. An algorithm, termed *training algorithm*, is used to adjust the weights. This process aims to minimize the difference between expected output and obtained output regarding the training data set. This type of learning is referred to as *supervised* learning. In mathematical terms, the learning of a neural network is the smooth interpolation of the data points constituting the input-output data sets. This is sketched in figure 3.3. Details on artificial neural networks can be found in the literature [26, 27].

In the context of performance evaluation the data used as training input stems from existing designs, and the performance of each of them must be known in advance to establish the evaluation model by means of supervised learning. In this context the neural network can be used to map the parameters of a design as input vector to the design performance as output. Provided this relation is adequately modelled by the neural network, when a new design is given as input to the model, i.e. a design that the network did not 'see' during the learning, the ANN provides the corresponding design performance at its output.

3.2.4 Shortcomings of data-driven approaches in design

An essential shortcoming of data-driven approaches for modelling design performance is the necessity to have sufficient training data available to establish the model. A large amount of data is necessary, so that any possible non-linear relation between input (design variables) and output (design performance) is accurately modelled. For a large number of variables generally a larger number of neurons and connections are necessary to model the input-output relation. That is, the complexity of the model needs to be high. The complexity is expressed as Vapnic-Chervonenkis (VC) dimension. With greater complexity the amount of training data samples required increases drastically for correct learning [26].

A design consists of a larger number of design variables that all, in one way or another, influence the design performance. A simple building may easily consist of tens of design objects, while each object has several parameters. This means the amount of input variables is generally high in architectural design. Therefore the amount of example cases required to establish an entirely data-driven model for performance evaluation is expected to be high in general, and may be excessively high.

Next to this is emphasized that in order to train the network, for every design case used as input to the model the corresponding performance should be known in advance, too. Obtaining a sufficient amount of reliable statements about the design performance is problematic due to the inherent complexity of the performance concept.

Also, the examples used for training should be well selected. This means the buildings from which the training data set is obtained should be representative of the design, for which the performance is to be measured. This makes it difficult to obtain the required large training data set. For example training data should stem from buildings of the same type, scale, style etc, and it may be difficult to find sufficient buildings that are representative enough of the project at hand. The difficulty is increased considering that building designs usually take place in specific contexts, where the relative importance among major performance aspects may vary strongly from project to project, even if a building subject to design is of the same type. This may be due to a client's unique preferences for instance that may be influenced by factors like economic situation or cultural context.

Another shortcoming of the data-driven approach is that knowledge available from experts regarding is minimally used in the model formation. Such knowledge concerns the details of the relations between design variables and design performance. Data-driven ANNs do not permit comprehension of the meaning of the parameters on the hidden layers of the model. This is because the parameters generally have no linguistic interpretation, so that an expert is unable to directly insert his/her knowledge into the model.

From the above consideration it is noted that the appropriateness of case-based approaches for performance evaluation in building design is questionable. The use of data-driven methodologies needs careful consideration. In particular their use should be concentrated to problems, where there is few expert knowledge available on an issue. Based on the above considerations it is concluded that performance evaluation, which exclusively relies on a data-driven approach, is unsuitable for architectural design.

3.2.5 Knowledge-driven approaches

Knowledge driven methodologies are based on inferring new facts from existing knowledge. That is, experts express their knowledge establishing a model. This is in contrast to the data-driven approach, where the knowledge model is formed using *data* from existing cases. One notes that generally speaking knowledge refers to information with strong semantics, while data is a numeric label associated with a physical event, and it is without semantics.

Next to symbolic systems of classic artificial intelligence (AI), most prominent knowledge-driven models are fuzzy systems, which belong to the computational intelligence (CI) methodologies.

It is easy to understand why classic AI methods are unsuitable for performance assessment of designs: they do not allow treating requirements that are satisfied to a *degree*. Classic propositional or predicate logic exclusively treat conditions that are either completely fulfilled or not at all. In performance evaluation the conditions refer to satisfaction of requirements. That is, classic logic exclusively handles *full satisfaction* or *zero satisfaction* of a requirement. However, design involves conflicting requirements, and therefore some requirements are generally partially fulfilled. As an example we may consider the conflicting nature of a requirement for low energy use of a building, while the costs for this building should be low at the same time. Increase of the performance of one aspect is at the cost of a decrease of the other one. Therefore, generally either requirement is satisfied to some degree situated somewhere between being completely satisfied and not satisfied at all.

Partial fulfilment of requirements can be handled using *fuzzy sets and fuzzy logic* [3] [28]. These concepts were introduced by Zadeh as generalization of classical crisp sets and traditional logic. In a recent communication[1] Zadeh suggested the following definition of Fuzzy Logic: *"Fuzzy logic is a precise system of reasoning, deduction and computation in which the objects of discourse and analysis are associated with information which is, or is allowed to be, imprecise, uncertain, incomplete, unreliable, partially true or partially possible. The importance of fuzzy logic derives from the fact that in much of the real world such information is the norm rather than exception. Fuzzy logic is designed to deal with problems which are avoided by traditional logical systems*." From this definition it is noted that the evaluation of the *suitability* of a design requires this kind of computation. This is because *suitability* or *performance* is essentially a linguistic concept, and as such it comprises from other linguistic concepts, such as *high functionality*, or *high sustainability*, etc. This means a highly abstract or general concept like *performance* is only very indirectly, related with physical quantities, such as sizes, distances, energy quanta, cost and time, namely via other linguistic concepts that are less abstract. This relationship will be further elucidated in the ensuing part.

In contrast to a classical set, a fuzzy set has set boundaries that are not sharp, but fuzzy. This means a certain object, may be a member of a set to a *degree*, and not merely *fully* or *not*. So, a fuzzy set is a class with a continuum of grades of membership, whereas ordinary sets have only two values expressing membership to the set, namely one or zero.

As an example let us consider the costs of a building. An expert, say a client governing the budget for the project, classifies any possible cost scenario *x* for the project. In this example the classification is by means of three fuzzy sets *A1*, *A2*, and *A3,* where set *A1* is labelled as *low cost, A2* as *moderate cost,* and *A3* as *high cost.* Each fuzzy set is characterized by a membership function $\mu_A(x)$. This function associates with each possible cost *x* a number in the interval [0, 1]. The number expresses the degree of membership of the cost *x* in the respective fuzzy set. It should be noted that the expressed membership functions are a precise, albeit subjective characterizations of their corresponding fuzzy sets.

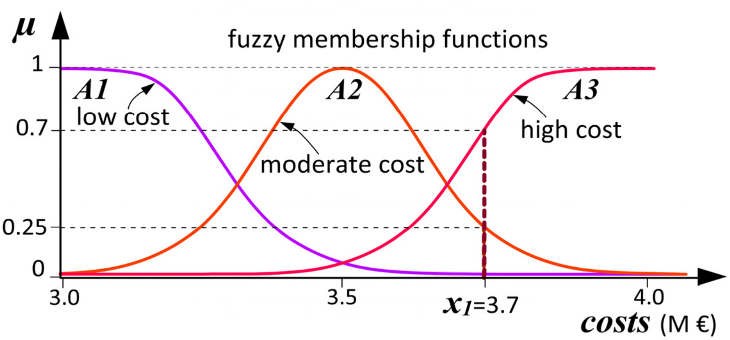


Fig. 3.4 Example of three fuzzy membership functions that characterize the costs

[1] Email communication addressing the members of the Berkeley Initiative in Soft Computing (BISC) group, November 2008

They may take different shapes to best represent an expert's knowledge on the classification concerned. An example of the fuzzy membership functions characterizing the fuzzy sets in the cost domain of a project is shown in figure 3.4.

To illustrate the meaning of the fuzzy membership concept let us take a design, where building costs are x_1 =3.7 million €. Then according to the client's fuzzy classification shown in figure 3.4 the costs x_1 are considered as *high costs* with a membership degree $\mu_{A3}(x_1)=0.7$ and at the same time they are considered as *moderate costs* with a membership degree $\mu_{A2}(x_1)=0.25$. This means the design concerned is more of a high-cost design than it is a design with moderate costs. Clearly, the design is hardly classified as a *low-cost* design according to the client's specification. This is reflected by the membership degree $\mu_{A1}(x_1)$ being close to zero. From this example one notes that in fuzzy set theory the same object may be associated simultaneously with multiple fuzzy sets within the same universe of discourse. It is also noted that the amount of sets used to characterize a universe of discourse defines a quality of the fuzzy classification termed *granulation*. The more membership functions subdivide the universe in fuzzy sets, the greater the granulation is, and visa versa.

As another example, in the same way as before, the size of the floor surface of a design may be classified into the fuzzy sets *small floor area (A1)*, *medium floor area (A2) and large floor area (A3)*. The corresponding membership functions are shown in figure 3.5. It is noted that in the figure for *A1* and *A3* two additional sets *A1'and* A3' are shown. These are are subsets of the sets *A1* and *A3* and labelled *very small area* and *very large area*. The attribute 'very' is termed concentrator and it is generally taken to be reflected by means of the squared membership function of the corresponding fuzzy set; i.e. $\mu_{A1'}(x)= \mu_{A1}^2(x)$ for very small, and $\mu_{A3'}(x)= \mu_{A3}^2(x)$ for very large. It is noted that the additional sets increase the granulation of our classification.

In the present case an example design with a floor area of 900m^2 has a membership of $\mu_{A1}(x_1)=0.85$ to the set of *small* floor area, $\mu_{A1'}(x_1)=0.60$ to *very small* area, and $\mu_{A1}(x_1)=0.25$ to medium area. This means the example design is primarily characterized as having a small surface area, while it resembles to a lesser extent to a design with a very small area. Clearly the area of *900m^2* is hardly considerable as a large area or very large area in the context of the present design. This is seen from the membership degrees for these sets, which are close to zero.

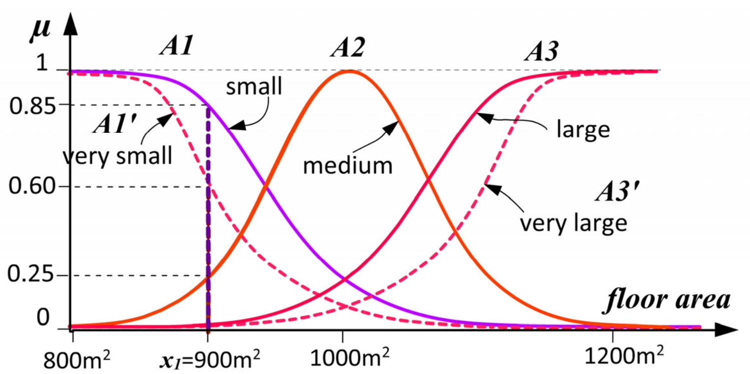


Fig. 3.5 Three fuzzy membership functions characterize the amount of floor area of a design

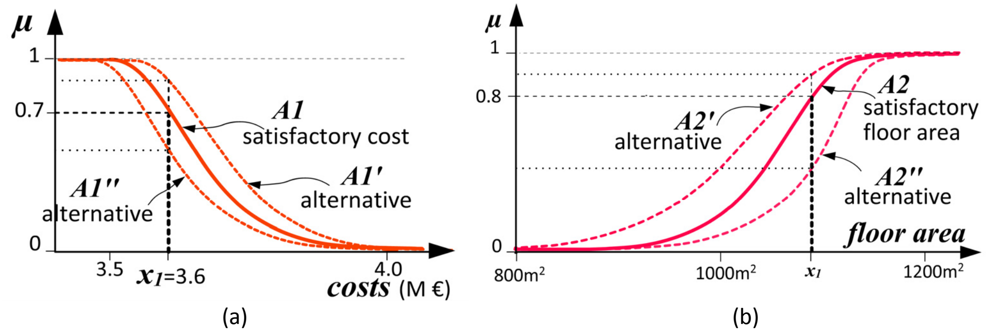


Fig. 3.6 Membership functions characterizing the fuzzy sets that express satisfactory cost for a building (a) and satisfactory size of its floor area (b)

3.2.6 Requirements expressed by means of fuzzy sets

Classifying design features regarding their membership in fuzzy sets opens avenues for advanced performance evaluation, where the partial satisfaction of requirements is taken into account: satisfaction of requirements can be expressed as a fuzzy set in a universe of discourse being a design variable. For example requirements on cost and floor area of a building may be expressed by the membership functions shown in figure 3.6. In the example moderate cost (or less than moderate costs), and a large floor area are required. Regarding the floor area, in this example the client is fully satisfied when the design has a floor area larger than *1170m²*. Then $\mu_{A2}(x_1)=1.0$. In this way any design is evaluated with respect to its degree of membership to the set of satisfactory costs and satisfactory floor area. Regarding costs, we can say the client is fully satisfied if the design costs less than 3.5 million €, i.e. in this case the membership to the set of satisfactory costs is $\mu_{A1}(x_1)=1.0$. As an example, in figure 3.6a a design is indicated, which costs x_1=3.6 million €. It has a membership degree of $\mu_{A1}(x_1)$=0.7 to the set A1 of satisfactory costs. So, the membership function determines how drastically satisfaction diminishes, when a solution deviates from the ideal condition.

As alternatives to the fuzzy set A1 expressing satisfactory building cost, two sets labelled A1′ and A1′′ are considered. Their fuzzy membership functions are shown in figure 3.7a. Set A1′ can be termed as a *relaxed* alternative to the requirement expressed by the set A1. This means for A1′ satisfaction diminishes less drastically when a solution differs from the ideal solution, compared to A1. In other words, a client expressing his/her cost requirement using A1′ instead of A1 can be said to be *more tolerant* regarding inaccurate attainment of what he/she considers ideal cost. This is seen from the example design costing x_1=3.6 million €: In case A1′ is used, the satisfaction is clearly higher compared to A1 being $\mu_{A1'}(x_1)$=0.85. In contrast to using A1′, a client using set A′′ as expression for satisfactory costs can clearly be considered as *more strict* regarding the attainment of ideal costs compared to using A1. This means satisfaction diminishes drastically for a design that costs more than 3.5 million €. Therefore the example design has a membership degree of only $\mu_{A1''}(x_1)$= 0.5, which is significantly less than when A1 is used. A1′can be said to express a requirement for a *quite* moderate cost, and A1′′ for *very* moderate cost.

In the same way two alternative fuzzy sets for the set A2 are considered and their membership functions are shown in figure 3.6b. A2′ expresses a *tolerant* requirement for a large floor area, whereas A2′′ expresses a *strict* requirement for a large floor area. This is reflected in the membership degrees for the alternative sets considering an example design. In figure 3.6b a design with a floor area x_1 of somewhat less than 1100m^2 has membership degrees of $\mu_{A2}(x_1)=0.8$ to the set of satisfactory floor area, while in the tolerant case $\mu_{A2'}(x_1)=0.9$, and in the strict case $\mu_{A2''}(x_1)=0.45$. So, clearly this design is less satisfactory when the strict requirement is applied.

The principle of representing requirements using fuzzy sets can be applied to other design requirements, ranging from functional requirements, perception related requirements, to energy consumption, and structural dimensioning requirements. In general the concept applies whenever there is some tolerance regarding full satisfaction of the respective requirement. It is emphasized that this is generally the case in design, due to the conflicting nature of requirements. It is noted that in the special case where there is no such tolerance for a certain requirement, a corresponding fuzzy set turns into an ordinary set with crisp set boundaries, manifesting that classical sets are a sub-class of fuzzy sets [29].

3.2.7 From fuzzy requirement satisfaction to performance

The ability to handle the graded satisfaction of design requirements is a unique feature of fuzzy sets. However, performance evaluation requires modelling the relation between the *entire* set of requirements and the design performance. That is, it should be modelled what the effect is on the total performance when multiple requirements are satisfied with respective degrees. This means we need a logic that can deal with partial satisfaction of several requirements, so that the consequence of their simultaneous partial fulfilment on performance is obtained. This logic is *fuzzy logic* [30] and a resulting model that relates the design variables to the design performance becomes a fuzzy system [4].

A fuzzy system is based on articulation of some commonsense knowledge by experts. The knowledge is expressed using plausible linguistic terms, creating fuzzy rules and numerically defined membership functions, such as the ones shown in figure 3.6a and 3.6b.

Fuzzy logic processing is also referred to as approximate reasoning. That means it is reasoning with fuzzy propositions. A fuzzy proposition is a proposition that contains a fuzzy membership function with associated fuzzy values. An example of a fuzzy proposition is: “Requirement A1 is satisfied.” Such a proposition may be *partially true*, and this is quantified by the membership degree μ for the fuzzy set representing the requirement satisfaction. This is in contrast to classic propositional or predicate logic, where a proposition is either true or false, but nothing in between.

In fuzzy logic, several fuzzy propositions are combined forming a fuzzy rule. Let us consider the following simple rule as an example: “If the costs are moderate (A1), and the floor area is high (A2), then the performance of the design (B) is high”, which consists of the three propositions *A1 is satisfied*, *A2 is satisfied*, and *performance B is high*. The first two propositions form the *premise* of the rule and the last one forms the *consequence* of the rule. In fuzzy logic terms such a rule can be expressed in generalized form as the *Zadeh-Mamdani fuzzy rule*

IF x_1 is A1 AND x_2 is A2 AND ... AND x_k is Ak THEN y is B,

Where in case of performance evaluation $x_1, x_2, \ldots, x_k$ are the design variables and y is the design performance; *A1, A2,...,Ak, B* are fuzzy sets defined over their respective universes of discourses *cost, floor area*, and *performance*, and characterized by their fuzzy membership functions $\mu_{A1}(x_1)$, $\mu_{A2}(x_2)$,..., $\mu_{Ak}(x_k)$, $\mu_B(y)$.

The fuzzy logic connective AND in the antecedent part of the rule is realized by taking the minimal membership function value, so that [31]

$$\mu_B(x_1, x_{2,\ldots,} x_k)=min\{\mu_{A1}(x_1), \mu_{A2}(x_2),\ldots, \mu_{Ak}(x_k)\} .$$

This means the fulfilment of the conditions in the rule has the consequence that output *B* is a fuzzy membership function. The membership degree of the resulting output in *B* is inferred to be the same as the minimum value of the membership function values in the premise part of the rule. This means the output membership function *B* is 'clipped off' at a height corresponding to the degree of truth of the rule's premise. The logic connective OR is realized by taking the *maximum* of the membership function values. We note that several rules can apply simultaneously and the combined effect of them is equivalent to an OR operation among the output membership functions.

One notes that as an alternative to the minimum operator for inferring AND operations, algebraic multiplication of the membership function values can be used scaling the output membership function B.

When the input value to a fuzzy system is crisp and the output value is expected to be crisp, too, then the output in the form of a membership function should be converted to a scalar number, which is referred to as defuzzification. Several techniques are used to accomplish this, e.g. determining the position of the center of gravity (COG) of the resulting membership function or the mean-of-maxima method (MOM). The MOM finds the value y' for the output variable y, which has maximum membership degree in the membership function B'. If there are several maxima the mean of their *y* values is taken. The COG is based on transforming a membership function B' is into a crisp value *y'* such that *y'* is the geometrical center of B'. It is accomplished by using [26]

$$y' = \sum \mu_B(v) \cdot v / \sum \mu_B(v) \ ,$$

where *v* are all the values from the universe *V* of the fuzzy variable *y*.

It is noteworthy to mention another major type of fuzzy rule introduced by Takagi and Sugeno [32], where the consequent part is a function of the input variables as opposed to a fuzzy set in case of the Zadeh-Mamdani type of fuzzy rule. The latter approach is particularly significant with respect to the conditional equivalence between fuzzy systems and neural systems.

3.2.8 Example of a fuzzy logic processing

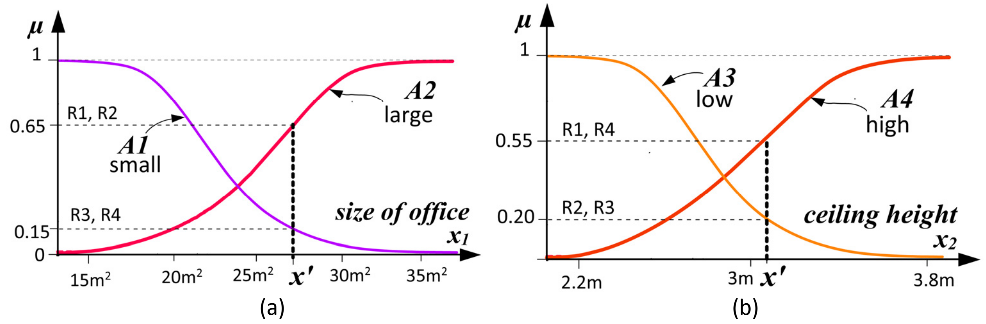


Fig. 3.7 Fuzzy sets used in the premise part of the four fuzzy rules

In order to illustrate fuzzy logic processing let us consider the following simple set of fuzzy rules expressing some performance related architectural knowledge:

R1: IF the office is large AND the ceiling is high, THEN spaciousness is high
R2: IF the office is large AND the ceiling is low, THEN spaciousness is medium
R3: IF the office is small AND the ceiling is low, THEN spaciousness is low
R4: IF the office is small AND the ceiling is high, THEN spaciousness is low

In fuzzy logic terms these rules are written as follows:

R1:	*IF*	x_1 *is A2*	*AND*	x_2 *is A4*	*THEN*	*y is B3*
R2:	*IF*	x_1 *is A2*	*AND*	x_2 *is A3*	*THEN*	*y is B2*
R3:	*IF*	x_1 *is A1*	*AND*	x_2 *is A3*	*THEN*	*y is B1*
R3:	*IF*	x_1 *is A1*	*AND*	x_2 *is A4*	*THEN*	*y is B1*

The membership functions that characterize the premises of the rules are shown in figure 3.7. The membership functions for the fuzzy sets *small office size (A1)*, and *large office size (A2)* are shown in figure 3.7a. The membership functions for the fuzzy sets *low ceiling (A3)* and *high ceiling height (A4)* are shown in figure 3.7b. The membership functions that characterize the consequent parts of the rules, are shown in figure 3.8.

Say we have an office ***x'*** with the following properties: it has a size of 27m^2 and a ceiling height of 3.0m. We ask: What is the *spaciousness y'* of the office according to the knowledge expressed via the four fuzzy rules *R1-R4*?

The resulting spaciousness of this office is about 0.7 on a scale between zero and one. This is seen in figure 3.8, where the centre of gravity (COG) of the membership function *B'* is located at the place *y'*=0.7 on the spaciousness dimension *y*. The membership function B' characterizes the spaciousness of the office x', and it is obtained as follows.

Let us consider rule number 1. For the office sized 27m^2, the fuzzy proposition forming the first part of the rule's premise, x_1 is A2, is fulfilled with a degree of $\mu_{A2}(x_1)$=0.65. The proposition forming the second part of the premise, x_2 is A4, is fulfilled with a degree of $\mu_{A4}(x_2)$=0.55. Executing the logical AND operation by applying the minimization rule, the consequent proposition, y is B3, has a membership degree of $\mu_{B'}(x_1, x_2)$=0.55 for this rule. This means the membership function B3 is clipped off beyond $\mu_{B'}$=0.55. This is indicated in figure 3.8 by means of the bold line segment labelled R1. In the same way the remaining three rules determine the shape of the consequent membership function B'. The final shape of B' is the maximization of the areas in the three fuzzy sets of spaciousness that are the consequences of the four rules' premises.

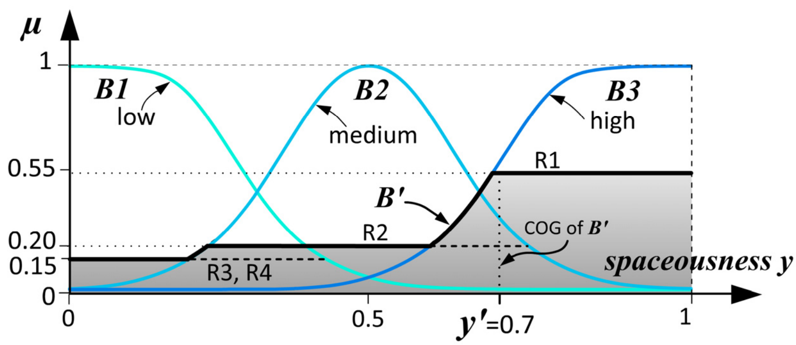


Fig. 3.8 Fuzzy sets of the consequence part of the four fuzzy rules

In the example above four rules mapped the two input variables into the consequent space. Designs have more than two variables, so the amount of rules needed to express the relations between design variables and the consequence on performance is large and it may be excessive. This inherent limitation of fuzzy systems is known as the curse of dimensionality [33]. The limitation refers in general to the problem that with increasing input variables to a fuzzy system the amount of fuzzy rules becomes excessive. This is especially problematic since an expert may be unable to articulate a large number of correct and consistent rules.

3.2.9 Neuro-fuzzy systems

To alleviate the limitations of fuzzy systems dealing with a large amount of input variables, neural network strategies were introduced into the systems in the 90s [5]. Enabling contribution in this respect was the establishment of *functional equivalence* between fuzzy systems and neural networks under certain lenient conditions [25, 34]. These conditions include that the shape of the network is a radial basis function (RBF) network having a Gaussian activation function, which plays the role of fuzzy membership function; the fuzzy inference method is multiplication; and the amount of fuzzy rules is equal to the amount of hidden layer nodes. The equivalence of fuzzy systems to neural networks by means of Gaussian radial basis function networks provides important insight into the symbiotic relation between fuzzy systems and neural networks. It permits the neural network training algorithms to be used in connection with fuzzy systems. This substantially circumvents the curse of dimensionality of fuzzy systems enabling them to deal with complexity.

This exciting symbiotic combination has found many soft computing applications and established its own technology. However, one should note that this is at the cost of a trade-off between transparency vs. complexity. That is, by introducing a neural network strategy into fuzzy systems the effects of the limitations of fuzzy systems are reduced at the cost of the systems' transparency. Transparency refers to the interpretability of the model. This means fuzzy systems gained in their ability to deal with complexity, at the cost of becoming 'grey' boxes that establish the relation between model input and output, say design variables and design performance, by means of learning from data samples. This is in contrast to the original spirit of fuzzy logic. Fuzzy logic was intended for computing with linguistic variables [28] and not for function approximation.

From the neural network side, the black-box feature of neural networks has turned to gray due to the transparency of the logic structure of fuzzy systems. The resulting systems are termed fuzzy-neural systems. They are hybrid systems where non-parametric neural systems are partially parameterized by fuzzy systems.

Although in the formal fuzzy-neural systems fuzzy a-priori-knowledge is used to pre-structure a neural network and thereby enhance its learning capabilities [26], such systems essentially belong to the data-driven approach. This means that the existing fuzzy neural networks perform *case-based reasoning* with its associated problematic issues.

The data-driven character of the existing neuro-fuzzy systems is an essential limitation for their application in design: available data from existing building cases is usually not enough in size, or not valid enough to yield reliable models for performance evaluation. This is due to the high dimensionality and conditionality of design requirements. Also, available expert knowledge cannot be inserted directly into the model when a data-driven approach is followed.

3.3. Fuzzy neural tree for performance evaluation

3.3.1 A novel knowledge-driven approach

The dilemma of formal neuro-fuzzy systems is that they are unable to maintain the transparency of the fuzzy logic, while dealing with complexity at the same time. Both features are in conflict with each. However, both are necessary for effective performance evaluation in design.

This dilemma may be alleviated using a special type of neural structure for a neuro-fuzzy system. This structure is known as *neural tree* [6, 7]. Neural trees are used to structure information. In this work neural trees are combined with the reasoning process of fuzzy logic. This way a special type of neural tree is obtained that is able to model design requirements. By integrating knowledge from experts directly into the tree structure the curse of dimensionality of a fuzzy logic system is circumvented, while the multitude of attribute relations and fuzziness of requirements are duly taken care of.

In the model both the neural network and fuzzy logic formalisms are considerably relaxed. By doing so, the gain obtained is twofold. On one hand, the ensuing network has more flexibility modelling the relation between elemental design requirements and design performance, compared to formal neuro-fuzzy systems. On the other hand, the *transparency* is achieved in the neural structure. This way knowledge is represented in comprehensible form throughout the network.

3.3.2 Fuzzy neural tree structures

Neural tree networks are in the paradigm of neural networks with marked similarities in their structures [35-38]. A neural tree is composed of one or several *root nodes* that are connected to *terminal nodes* via *internal nodes*. The terminal nodes are also termed *leaf nodes*. The connection is established via internal nodes that 'branch out' from the root nodes in a tree like manner. A neural tree-structure is seen in figure 3.9. Connections between two nodes in the tree have a weight attributed to them. The weights quantify the "strength" of the relation. The non-terminal nodes represent neural units. The neuron type is an element introducing a non-linearity simulating a neuronal activity. In the present case, this element is a Gaussian function. The Gaussian has several desirable features for the goals of the present study; namely, it is a radial basis function ensuring a solution and it has smoothness. At the same time it plays the role of a fuzzy membership function in the tree structure. Namely the tree structure is considered to be a fuzzy logic system, as its outcome is based on fuzzy logic operations and associated fuzzy reasoning. Therefore the tree is named as *fuzzy neural tree*.

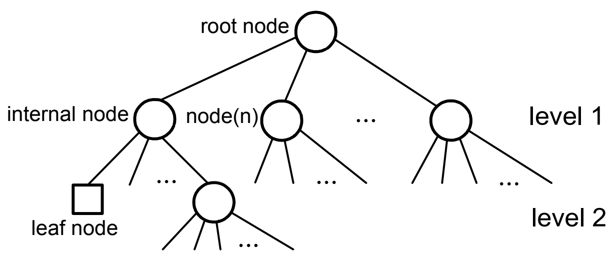


Fig. 3.9 The structure of a neural tree

The neural tree is used for performance evaluation by structuring the relations among aspects of performance. This is seen from figure 3.10 where a generic neural tree structure for performance evaluation is shown. The root node takes the meaning of *high design performance* and the inner nodes one level below are the aspects of the performance. The meaning of each of these aspects may vary from design project to project and it is determined by experts. They take meaning such as *high functionality*, *high sustainability*, *low costs*, and *high perception value*. One level further below is formed by the constituents of these aspects, which are termed sub-aspects. For example considering the aspect of sustainability, sub-aspects may be *low energy use*, *spatial flexibility*, and *high recyclability*. These sub-aspects may be further unraveled into their aspects. Provided there is sufficient knowledge at hand, the tree may comprise a large amount of sub-aspect levels that further refine the performance evaluation model. The nodes at the lowest level of the tree form the terminal level. At this level the elemental design requirements are modeled. As an example let us consider that the aspect one level above some elemental requirements is *high perception of spaciousness*. In this case elemental requirements are for example large *surface area* and *high ceiling*. Such requirements at the terminal level may refer to physical quantities in a design, and they are in the form of a fuzzy set. Ideally the input values are automatically measureable from a design, because then the neural tree model can be employed in a systematic search for maximal performance.

Aspects generally differ regarding their relative influence on the performance at the root node. In order to specify the relevance of an aspect for the concept at one level higher in the tree, experts assign a weight to the respective connection. This is shown in figure 3.10 by means of the variables w_i.

Preparing a neural tree in this way, i.e. by specifying the structure and the connection weights among the nodes of the neural tree, one or several designers determine how design performance is evaluated. That is, they define what the constituents of the performance are. They also specify the relative importance among the constituents using their knowledge.

When the tree is used for performance evaluation the root node of the tree computes the design performance. The major aspects determining the performance are computed at the nodes one level below the root node. Sub-aspects of the main aspects are computed one level further below, and so on. This means during evaluation of a design alternative the tree is stimulated at its leaf nodes involving a fuzzification process. This process is the mapping of the value of a design variable to a degree of satisfaction of an elemental requirement. The resulting information is then processed by the inner nodes in the form of AND operations. The principle is that the satisfaction of an elemental requirement contributes to the satisfaction of the performance aspect at the next higher level in the tree hierarchy; finally the satisfaction contributes to the total design performance computed at the root node. In this process the relative degree of contribution is in accordance with the relevancies specified by the experts.

Although the principle is apparently simple, the computations performed at the nodes involve multi-dimensional Gaussian membership functions. The shape of the function in every dimension depends on the input weights coming to the node. Therefore, the entire operation of the model, even when a tree consists of only a few inner nodes, is not simple.

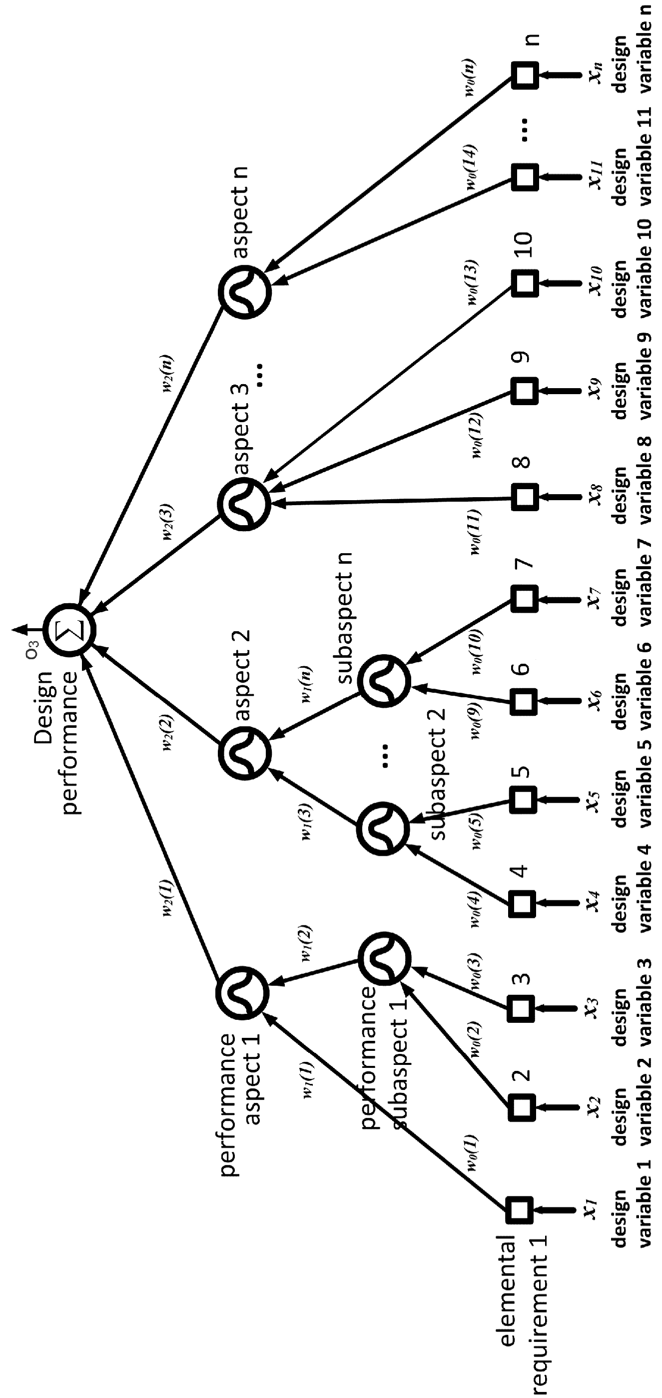


Fig. 3.10 Generic fuzzy neural tree structure for performance assessment

3.3.3 Fuzzy logic processing at the nodes

In a neural tree, the root node is an output unit and the terminal nodes are input units. These nodes are connected via internal nodes. Each terminal node is labelled with an element from the terminal set $T=\{x_1, x_2,...,x_n\}$, where x_i is the *i*-th component of the external input vector *X.* This is illustrated in figure 3.11. Each link *(i,j)* represents a directed connection from node *i* to node *j.* A value w_{ij} is associated with each link.

The node outputs of the neural tree are computed in the same way as computed in a feed-forward neural network. In this way, neural trees can represent a broad class of feed-forward networks that have irregular connectivity and non-strictly layered structures. In particular, in the present work the nodes are similar to those used in a radial basis functions network with the Gaussian basis functions.

In the neural tree considered in this work an inner node processes the input coming to it by means of a Gaussian radial basis function [3]. The output is obtained by

$$f(X) = w\phi(\|X-c\|^2) \tag{3.3}$$

where $\phi(.)$ denotes the Gaussian basis function, ***c*** denotes the centre of the basis function, and *X* denotes the input to the node. The Gaussian is used in this model due to its relevance to fuzzy-logic. In particular use of the Gaussian function ensures equivalence between fuzzy systems and neural networks [25], so that the neural computations being described are equivalent to fuzzy logic operation. This equivalence is rooted in the fundamental isomorphic property of the exponential function. That is, multiplication of two Gaussians yields again a Gaussian function, which differs from its factors only with respect to the function parameters. This property of the Gaussian forms the link between the *symbolic* paradigm represented by fuzzy modelling, and the sub-*symbolic* paradigm represented by neural network modelling [26]. The isomorphic property of the Gaussian function is the key for integration of linguistically structured expert knowledge into radial-basis-function neural networks (RBF) networks as it is exercised here.

The output of the *i*-th terminal node is denoted w_i and it is introduced to an inner node. One notes that w_i represents the degree of satisfaction of the *i*-th requirement.

That is, the membership degree in the fuzzy set defined at the terminal node representing an elemental requirement. The output w_i of terminal node *i* is related to each component X_i of the input vector *X* coming to inner node *j* by [3]

$$X_i = w_{ij} w_i \tag{3.4}$$

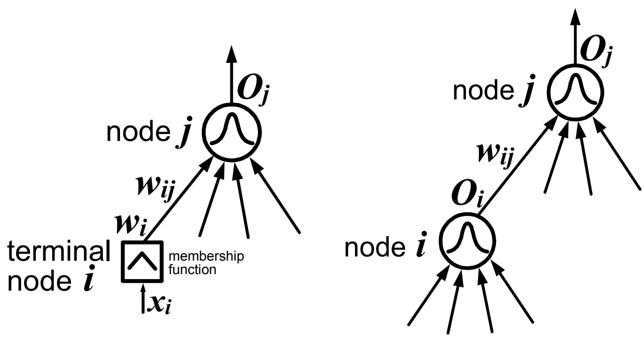


Fig. 3.11 Details of a neural tree structure showing the different type of node connections

where w_{ij} is the weight assigned to the connection between terminal node *i and* inner node *j* by the experts that formed the tree structure. This means the output is scaled by the weight characterizing the significance of the node.

The centres of the basis functions are set to be the same as the weights of the connections arriving at that node. Therefore, for a *terminal node* connected to an *inner node*, we can express the inner node output denoted by O_j, as [3]

$$O_j = \exp(-\frac{1}{2}\sum_i^n \left[\frac{X - w_{ij}}{\sigma_j}\right]^2), \tag{3.5}$$

which becomes due to Eq. (3.4)

$$O_j = \exp(-\frac{1}{2}\sum_i^n \left[\frac{w_{ij}(w_i - 1)}{\sigma_j}\right]^2) \tag{3.6}$$

where *j* is the number of the node; *i* denotes consecutive numbers associated to each input of the inner node; *n* denotes the highest number of the inputs arriving at node *j*; w_i denotes the degree of membership being the output of the *i-th* terminal node; w_{ij} is the weight associated with the connection between the *i*-th terminal node and the inner node *j*; and σ_j denotes the width of the Gaussian of node *j*. The width of the Gaussian function is used to measure the uncertainty associated with the node inputs designated as external input *X*. One notes that the fuzzy logic operation performed at each node is an AND operation among the input components X_i coming to the node. This is realized by applying the multiplication rule for the AND operation. The multiplication equals summation at the exponent of the Gaussian function in Eqs. (3.5) and (3.6).

In the same way as in Eq. (3.4), for an inner node connected to an inner node the output O_i of inner node *i* is related to each component X_i of the input vector *X* coming to node *j* as given by [3]

$$X_i = O_i \, w_{ij}, \tag{3.7}$$

where w_{ij} is the weight determined by expert judgement characterizing the connection between inner node *i* to the inner node *j*. The relation is illustrated in figure 3.11. Due to Eq. (3.7) Eq. (3.5) becomes

$$O_j = \exp(-\frac{1}{2}\sum_i^n \left[\frac{O_i w_{ij} - w_{ij}}{\sigma_j}\right]^2), \tag{3.8}$$

which can be written as

$$O_j = \exp(-\frac{1}{2}\sum_i^n \left[\frac{w_{ij}(O_i - 1)}{\sigma_j}\right]^2). \tag{3.9}$$

We can express Eqs. (3.6) and (3.9) as follows

$$O_j = \exp(-\frac{1}{2}\sum_i^n \left[\frac{(w_i - 1)}{\sigma_{wj}}\right]^2) = \exp(-\frac{1}{2\sigma_j^2}\sum_i^n \left[\frac{(w_i - 1)}{1/w_{ij}}\right]^2) \tag{3.10}$$

$$O_j = \exp(-\frac{1}{2}\sum_i^n \left[\frac{(O_i - 1)}{\sigma_{wj}}\right]^2) = \exp(-\frac{1}{2\sigma_j^2}\sum_i^n \left[\frac{(O_i - 1)}{1/w_{ij}}\right]^2), \tag{3.11}$$

where

$$\sigma_{wj} = \frac{\sigma_j}{w_{ij}}. \tag{3.12}$$

Referring to (3.10)-(3.12), the following observations are essential:

- For each node, parsimoniously, a common width is defined in place of several ones, each for a term in the summation.
- The centre of a Gaussian is always equal to one.
- The width of a Gaussian is scaled by the input weight w_{ij}. In other words, as to width, the shape of Gaussian fuzzy membership function is dependent on the input weights w_{ij} of the respective Gaussian.
- The derivative with respect to w_i for the terminal nodes, and with respect to O_i for non-terminal nodes is always positive. Namely, for node input coming from terminal nodes, from Eq. (3.10)

$$\frac{\partial O_j}{\partial w_i} = -\frac{{w_{ij}}^2}{\sigma_j^2}(w_i - 1)O_j \tag{3.13}$$

Since $w_i \leq 1$ and O_j is positive, it follows that

$$\frac{\partial O_j}{\partial w_i} \geq 0 \tag{3.14}$$

In the same way, for node input coming from other inner nodes, from Eq. (3.11)

$$\frac{\partial O_j}{\partial O_i} = -\frac{{w_{ij}}^2}{\sigma_j^2}(O_i - 1)O_j \tag{3.15}$$

Since $O_i \leq 1$ and O_j is positive, it follows that

$$\frac{\partial O_j}{\partial O_i} \geq 0 \tag{3.16}$$

It is emphasized that the centre location of a Gaussian at an inner node with inputs from non-terminal nodes is always located at the point w_i=1, respectively O_i=1, as this is also seen from figure 3.12. This is a novel type of neural computation, which is slightly different from conventional radial basis function (RBF) type computation. In the latter neural computation the centres are determined by other means, clustering for instance. Placing all centres at c=1 circumvents the curse of dimensionality of fuzzy systems. This is because the RBF centre of each node is determined independently from other nodes.

One notes that the particularity of the fuzzy neural tree given by Eqs. (3.14) and (3.16) implies that with increasing satisfaction of an elemental requirement the total design performance increases as well. In the fuzzy logic terminology, approaching the maximum of the fuzzy membership function at the input is reflected at the output of the model. This means the neural tree output is an increasing function of the inputs. In other words, with the increasing values of the input, the tree output is always increasing and vice versa.

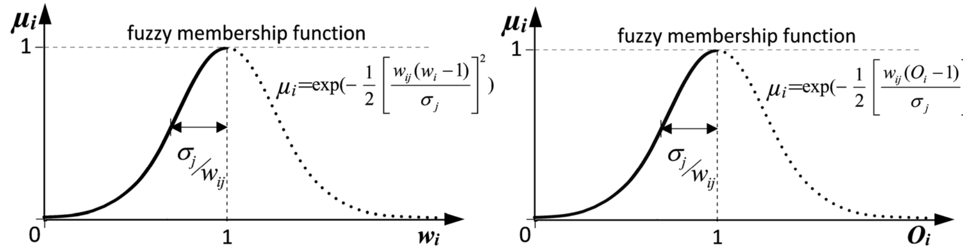


Fig. 3.12 The fuzzy membership functions at inner nodes

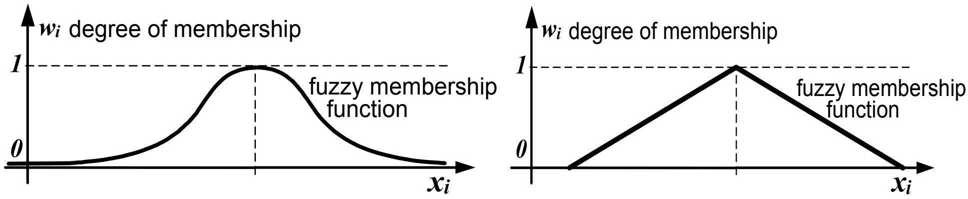


Fig. 3.13 Two possible fuzzy membership function type among many others, at the terminal node.

This is consistent with performance evaluation in architectural design. Namely when a certain requirement is satisfied more it will increase the overall performance. This is valid even though the increase may be slight due to insignificance of the requirement in the context of other requirements.

This property also implies that in the fuzzy logic processing only the left side of the Gaussian membership functions of the inner nodes are used. This is shown in figure 3.12, where the right side of the Gaussian is represented with a dotted line. All the input weights of the nodes are dependent on the neural tree structure and determined by the domain knowledge. At the terminal nodes membership functions are not necessarily Gaussian; they can be triangular, among many other types, depending on the elemental requirement being modelled. Some possible membership function types at the terminal node are illustrated in figures 3.13. It is noted that that the degree of membership, which quantifies the degree of satisfaction of the respective requirement, is denoted by w_i in figure 3.13.

In the neural tree structure, only the root node performs a simple weighted summation of the inputs coming from the immediate layer below. Terminologically, this is the de-fuzzification process for the final outcome. Therefore the weight connected to the root node sums up to one.

3.3.4 Imposing the condition of consistency

The limitations of fuzzy logic systems are due to the imprecision of (i) the number of membership functions, (ii) the membership function shape, (iii) the location of a membership function, and (iv) the curse of dimensionality. Having an expert express his/her performance related knowledge in the form of the particular fuzzy neural structure described above, sources (i), (iii), and (iv) are eliminated. Concerning the shape of the membership functions it is selected as Gaussian, so the sole remaining issue is determination of the widths of the Gaussian functions at the inner nodes. The determination of the widths is performed by the imposition of what is termed as *consistency conditions* in this framework. The consistency conditions serve as boundary conditions for the neural tree model.

In figure 3.1 the neural tree nodes play the role of information processors that perform fuzzy AND operations. It is noteworthy to point-out that the input information to a node first is fuzzified by means of a Gaussian membership functions, thereafter AND operation is performed, to this information. The fuzzification is accomplished being directly related to the associated input weight. The weights are domain knowledge and they sum up to unity. This means the knowledge provided to the neural tree is directly used together with the input information, so that the commensurate outputs corresponding to all inputs are obtained with fuzzy AND operation. This is a novel way of performing AND operation in the sense that fuzzy numbers are *directly* obtained from the fuzzy membership functions and they are directly multiplied as an arithmetic operation, where directly means the whole

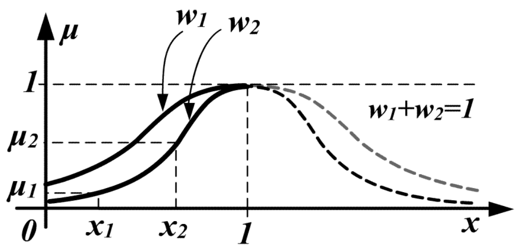


Fig. 3.14 Input fuzzification

process takes place inside the node, without explicit fuzzy set operations. This is explained schematically in figure 3.14. In this figure only two inputs are considered without loss of generality. The variables w_1 and w_2 are input weights determining the width of the Gaussian membership function. It is noted that in the figure $w_1<w_2$. This is clear from Eqs. (3.10) and (3.11).

For two inputs, two distinct standard deviations are defined. In particular, the inputs can be equal, i.e., $x_1=x_2$. This particular case occurs when the outputs of the two nodes delivering the inputs to the node we are considering are equal. This case is illustrated in figure 3.15a. In figure 3.15b it is clear that, if w_1 and w_2 are equal then the AND operation is expressed by means of a single Gaussian.

Since σ_j is a free parameter, by giving an appropriate width via Eqs. (3.10) and (3.11), the result of the AND operation is given by

$$
\begin{aligned}
&\mu = \mu_1 = \mu_2 \\
&x = x_1 = x_2 \qquad (3.17)\\
&O_i = \mu_1 \cdot \mu_2
\end{aligned}
$$

This is illustrated in figure 3.16 , where the left part of the Gaussian is approximated by a straight line.

In figure 3.16, optimizing the σ_j parameter, we obtain

$$O_j \cong x \qquad (3.18)$$

for all values of x and O_j between zero and one. In any case, for a node in the neural tree, Eq. (3.18) is satisfied for $x=O_j=0$ (approximately) and for $x=O_j=1$ (exact) inherently, while μ_1 and μ_2 are increasing function of x_1 and x_2. Therefore a linear relationship between O_j and x in the range between 0 and 1 is a first choice from the fuzzy logic viewpoint; namely, as to the AND operation at the respective node, if inputs are equal, that is $x=x_1=x_2$ then the output of the node of x_1 AND x_2 is determined by the respective *triangular membership*

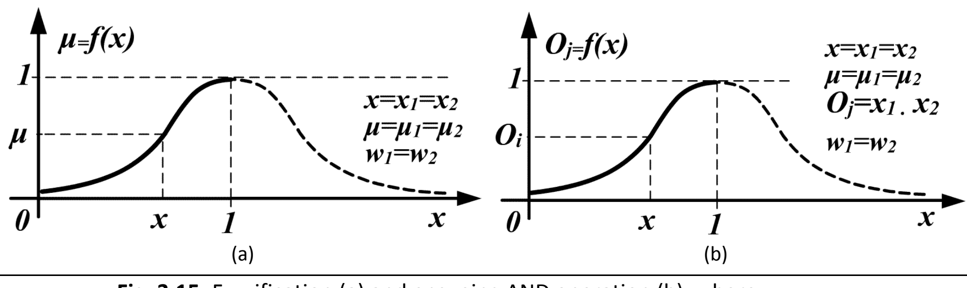


Fig. 3.15 Fuzzification (a) and ensueing AND operation (b), where $w=w_1 = w_2$

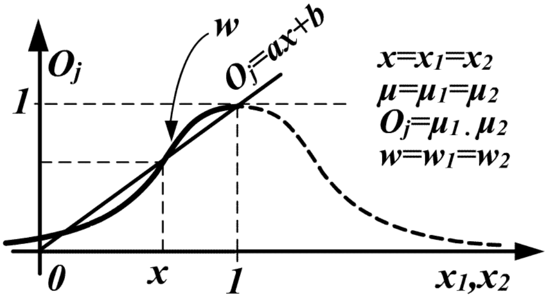


Fig. 3.16 Linear approximation to Gaussian function

functions in the antecedent space. Triangular fuzzy membership functions are the most prominent type membership functions in fuzzy logic applications. These membership functions are represented by the data sets in table 3.1. In general, the data sets given in the table are named in this work as '*consistency conditions*'. They are used to calibrate the membership function parameter σ. This is accomplished by optimization.

The consistency data sets are shown in Table 3.1 as input and output applied to the tree for a supervised training, in the terminology of neural networks. The input consists of values for $x_1, x_2,\ldots,x_n$ and the output is the resulting performance value at the root node denoted by $f(x_1, x_{2,\ldots,} x_n)$ in the table. The latter value represents the case for ideal consistency.

As result of the training, all the Gaussian widths are determined, so that the consistency conditions are established for the tree. It is emphasized that if all the inputs in Eqs. (3.10) and (3.11) are unity, then the output at the root node is also unity. Therefore this input-output pair is not included in the training set. Namely, for the input $x_1=1, x_2=1,\ldots,x_n=1$ to an inner node the output of the node consequently also becomes one; This is seen from Eq. (3.10), where the centres of the basis functions are given by a vector $\boldsymbol{c}=\{1, 1, 1,\ldots.,1\}$, that is $c_i=1$. This implies that the Gaussian fuzzy membership functions have their maximum value at the point where all inputs are one. When the outputs of the terminal nodes are one, any connected inner node will also provide one as its output value. Consequently the weighted summation at the root node will also yield one. This is because the weights of the connections between penultimate nodes and root node sum up to one. In other words, for the ideal condition that all design requirements are perfectly satisfied the design performance also takes the maximum value, as we should expect.

Table 3.1 The two datasets to establish the consistency conditions

	x_1	x_2	x_3	…	x_n	f(X)
1	.1	.1	.1	.1	.1	.1
2	.2	.2	.2	.2	.2	.2
3	.3	.3	.3	.3	.3	.3
4	.4	.4	.4	.4	.4	.4
5	.5	.5	.5	.5	.5	.5
6	.6	.6	.6	.6	.6	.6
7	.7	.7	.7	.7	.7	.7
8	.8	.8	.8	.8	.8	.8
9	.9	.9	.9	.9	.9	.9

It is emphasized that the neural tree output follows the trend of input w_i representing the associated degree of membership. Considering this property, the *consistency conditions* refer to the fact that if all the inputs w_i are a certain value, the system output at the root node should also take this very value. This means, if the design requirements are poorly satisfied the design performance is low, too. If the requirements are moderately satisfied the performance is moderate, and if they are highly satisfied performance is high. These conditions are inherent to the concept of performance.

Imposition of consistency in this way is a marked difference between a neural network and a neural tree. Namely, in neural networks the weights are determined by learning and the network is a *universal approximator*. In the fuzzy neural tree weights are determined by domain knowledge, so that it is not a universal approximator. It is a *soft computing estimator*, estimating the reflection of the input composition to the output with the presence of given structural relationships.

At this point a few observations are due, as follows. If a weight w_{ij} is zero, this means the significance of the input is zero, consequently the associated input has no effect on the node output and thus also the system output. Conversely, if a w_{ij} is unity, this means the significance of the input is highest among the competitive weights directed to the same node. This means the value of the associated input is extremely important and a small change about this value has big impact on the node output O_j. If a weight w_{ij} is somewhere between zero and one, then the associated input value has some possible effect on the node output determined by the respective AND operation via Eqs. (3.10) and (3.11). In this way, the domain knowledge is integrated into the logic operations.
The general properties of the present neural tree structure re as follows.

- If an input of a node is small (i.e., close to zero) and the weight w_{ij} is high, then, the output of the node is also small complying with the AND operation
- If a weight w_{ij} is low the associated input cannot have significant effect on the node output. This means, quite naturally, such inputs can be ignored.
- If all input values coming to a node are high (i.e.,close to unity), the output of the node is also high complying with the AND operation

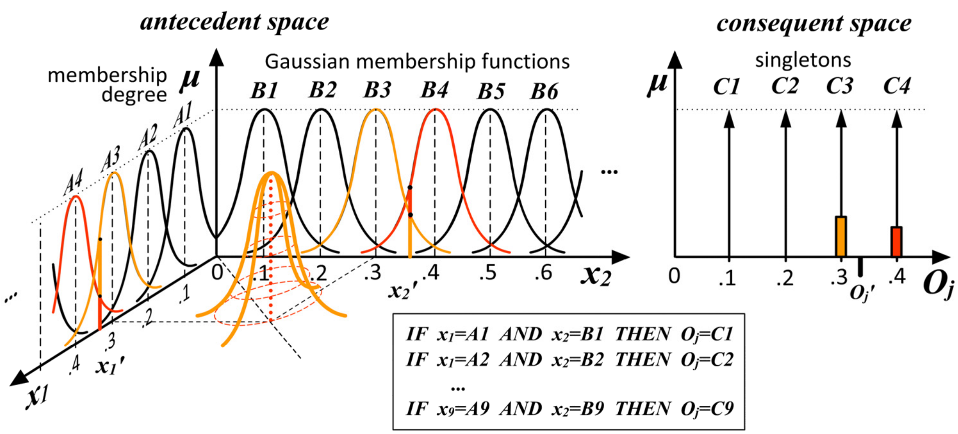


Fig. 3.17 Illustration of fuzzy logic information processing at an individual neural tree node with two inputs

- If a weight w_{ij} is high the associated input x_i can have significant effect on the node output.

It might be of value to point out that, the AND operation in a neural-tree node is executed in fuzzy logic terms and the associated connection weights play an important role on the effectiveness of this operation.

The fuzzy logic interpretation of the nodal computation is illustrated in figure 3.17. From the figure it is seen that the input x_1 and x_2 yield membership degrees at several membership functions in the antecedent space. After multiplication of these membership degrees, the membership degrees in the consequent space are determined at singleton membership functions, where every singleton belongs to the rule involving the corresponding Gaussian membership functions. The fuzzy rules are written in the figure as well.

It is emphasized that by means of the above described approach, the locations of the Gaussian membership functions at the non-terminal nodes are well-defined. The shape of the fuzzy membership functions at the non-terminal nodes are Gaussians due to logic operations. Namely, each input to a node has contribution to the output of that node based on a logic AND operation, which is realized through multiplication of the Gaussians. The centre location of the *i*-th Gaussian membership function is selected as w_{ij} , so that certain features described above are maintained. In contrast to the case in neural networks, here the weights are determined by domain knowledge, whereas in neural networks they are determined by training.

The assignment of the weights is accomplished as follows. If there are several input weights that belong to a certain inner node, these weights are determined with respect to each other. This means the relative degree of significance the input has for the node are determined. The inputs can all be pair-wise compared and normalized afterwards for convenience. After normalization, if some weights are not relatively significant, they can be discarded. However, normalization is not necessary, as this will become clear below.

3.3.5 Performance evaluation of designs

The inner node performs an AND operations using Gaussian membership functions. This way the requirement satisfactions determined at the lower level propagate through the neural structure. Only the root node performs a simple weighted summation of the inputs coming from the immediate layer below providing the design performance value on a scale between zero and unity. In terms of fuzzy logic, the latter computation is the de-fuzzification process for the final outcome.

As an example let us consider that a neural tree represents the performance of an office space. If all the required features of the space are highly satisfactory, then the performance of the office is high, too. In fuzzy logic terminology, if all the membership degrees that specify the office in the sets of satisfied requirements are close to one, then the neural tree output is expected to be close to one, too. If for example the size of the office is large, the indoor climate is comfortable, the view is attractive, and the privacy is high, then the office space is a good one.

Due to the increasing function property of the tree any increasing satisfaction of any requirement will be reflected as an increase of the design performance at the output. However, the *degree* of performance increase is still dependent on the tree structure representing the domain knowledge established.

Due to the conflicting nature of design requirements, the degree of satisfaction among requirements generally varies from requirement to requirement. This means some of the requirements are poorly satisfied, while others are highly satisfied, and others are moderately satisfied. The total performance is obtained with precision using the fuzzy neural tree. This means through the tree the performance-related knowledge is simultaneously treated on the detail level of elemental satisfaction, as well as on the 'higher' level of complex concepts, such as *functionality*. Such treatment across levels of abstraction is difficult to realize using conventional means. This is apparently due to the limitation of the human mind to resolve detail and store information [14].

Considering human memory limitations [18, 19, 39] it is easy to understand that a decision maker has difficulty to pay attention to the total performance of the design, and at the same time consider the diverse and concrete details. These issues are usually treated in sequence [40, 41]. In general attention has to be spread out among these different levels of consideration, whereas it would be desirable to have full awareness on every level at the same time, to ensure deepest insight into the matter. As soon as we shift our attention among issues, due to the elusive character of biological memory we cannot guarantee that we hold our previous consideration in memory well enough as to ensure the current one is adequately combined with the previous one. The knowledge model shown is a leap forward in this respect. It aids architects to express their knowledge on various levels of generalisation using their full attention for every aspect.

3.3.6 Example of a fuzzy neural tree for performance evaluation

An example of a neural tree structure for performance evaluation is shown in figure 3.18. The output of the model is designated as performance of a design. It concerns positioning of some buildings, so that a number of functional and perceptual requirements are satisfied.

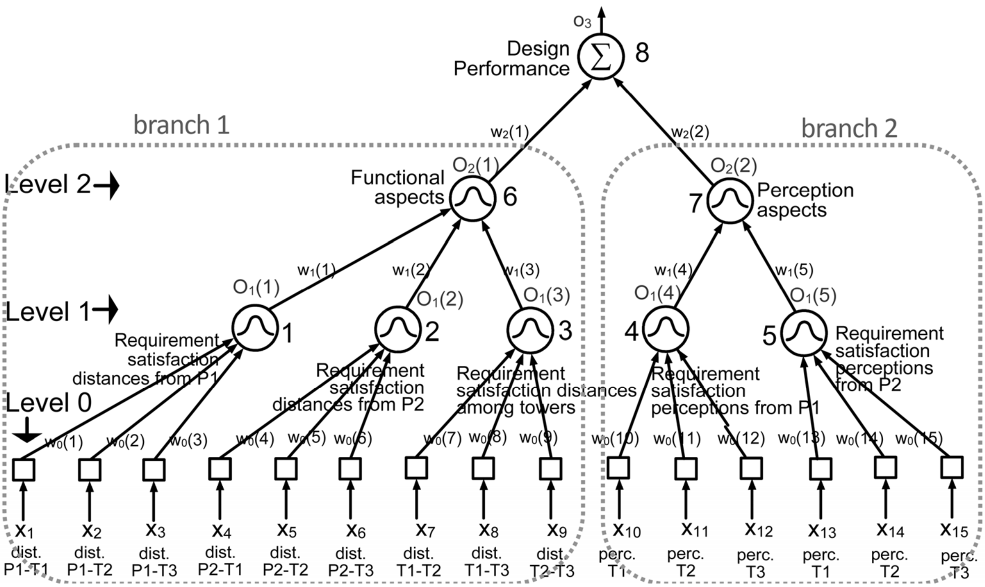


Fig. 3.18 An example of a neural tree employed for design

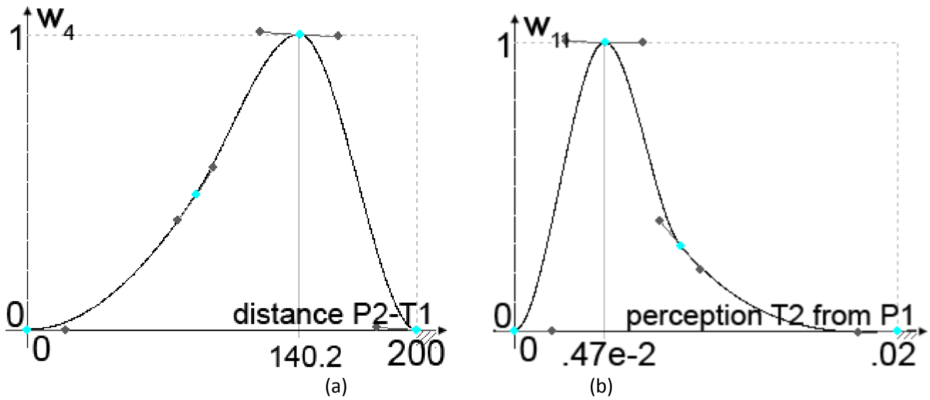


Fig. 3.19 Membership functions at the terminal nodes.

There are two options for consistency training. In one option, if all the inputs to the tree are uncorrelated and the tree is in the form of several branches, then the training is performed for each branch. A branch is defined as an independent set of nodes connected to the root node. The output of a branch is taken as the node one below the root node. For example in figure 3.14 there are two branches. The first branch has node number 6 as output node. The nodes connected directly or indirectly to this node form the first branch. The second branch has node number 7 as output node, and the nodes connected directly or indirectly to this node form the second branch. When the consistency conditions are established for all branches, it includes inherently the root node also.

As another option, if the inputs to the tree are correlated, then to allow for this correlation the training establishing the consistency conditions includes also the root node. This means for the whole tree structure one common training is performed. It is noted that after consistency training only the relative values of the weights coming to a node are in effect. This indicates the effectiveness of each weight on the node relative to each other. In essence, such relativity is taken care of by the Gaussian widths that are established in the training.

After the consistency conditions are established by training, for the inputs at the terminal nodes appropriate fuzzy membership functions are used. These are based on domain knowledge. As examples of fuzzy membership functions at the terminal level of the tree in figure 3.18 two functions are shown in figure 3.19. From figure 3.19a it is seen that the requirement w_4 is maximally satisfied when the design variable, in this case a distance between two objects, takes a value of 140m. In case the distance is larger the satisfaction reduces drastically, whereas for shorter distances the satisfaction diminishes less drastically. In the same way figure 3.19b is interpreted, however the input variable concerns a degree of perception in this case.

The employment of fuzzy membership functions has two implications. The neural tree accepts input between zero and one as a dimensionless quantity. This value is the degree of requirement satisfaction at the respective terminal node. Therefore, during the computations we are not concerned with units and all the node outputs are also between zero and one. This makes interpretation of the results easy, not only at the root node but also throughout the tree structure.

Another implication worth to mention is the circumvention of standardization of the inputs. Namely, since the inputs to the tree are some physical quantities, in general they are all measured with their respective units and there can be a big difference among these units. This is termed as “stiffness” in science. Since larger values at the input produce larger outputs, it would not be fair to directly apply the measurement values with their respective units. By means of the membership functions all inputs are ‘standardized’ in some way similar to the standardization of statistical data according to a Gaussian distribution, in the statistical analysis. By doing so, the ‘stiffness’ is circumvented.

As a recapitulation of the previous sections we note that the novel neuro-fuzzy system allows evaluating design performance using domain knowledge that is integrated into the system. Therefore it is a knowledge-driven modelling, where the fuzzy modelling concept is essential.

3.3.7 Abstraction and generalization in performance models

Abstraction is a general concept in science. It is of utmost relevance because it is the essential mechanism when we establish a model of a physical phenomenon. An entity can be either physical or abstract. Abstract entities have no physical existence. They are subject to comprehension. In contrast to that physical entities exist in reality, so that they can be subject to measurement. For example a circle is an abstract entity as its description is entirely by means of abstract terms, i.e. mathematical terms, whereas a wheel is a physical entity. Abstract entities generally denote a physical entity.

It is emphasized that the input to the tree is *fuzzified* by determining its degree of membership in a fuzzy set that represents the satisfaction of a requirement. The membership degree is used in the ensuing logic computations to obtain the design performance. This fuzzification process is a mapping of the quantity of the design variable from the *physical* domain into the *abstract* domain. Any ensuing logic operation is in the abstract domain. Within the abstract domain we may distinguish *levels* of generalization. As we are moving away from the level of specific elemental requirements towards concepts with higher degrees of complexity, the level of generalization becomes higher. The design performance is the most general concept involved. It consists of less general/more specific concepts, and they become more and more specific towards the input level of the neural model. Finally the least general level is the input level to the tree consisting of the terminal nodes. This is illustrated in figure 3.20.

From this viewpoint fuzzy neural trees (FNTs) perform approximate reasoning that involves abstract concepts with different levels of generalization. This means such a model evaluates the performance of alternative designs in a human-like manner.

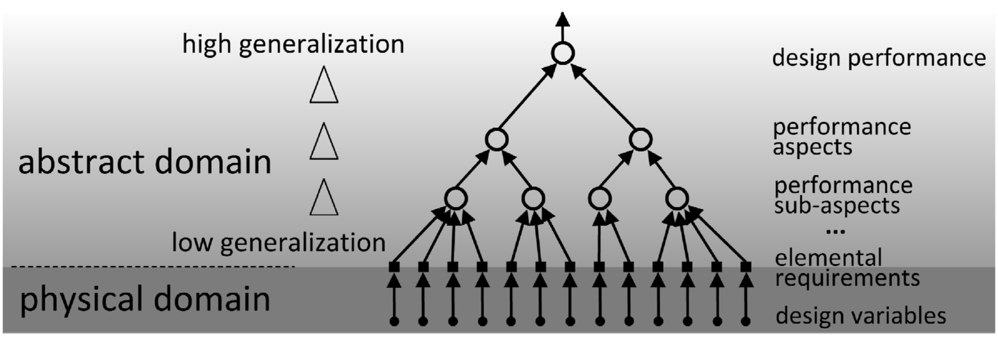


Fig. 3.20 Degrees of abstraction in performance evaluation using a fuzzy neural tree

3.4. Discussion

To clarify how the neural structure and the weights are determined in the present approach it is to be noted that the procedure of this chapter is a novel approach to neuro fuzzy modelling. In particular the model does not handle *data*, in contrast to the existing neuro-fuzzy modelling approaches.

It models commonsense knowledge, remaining in the paradigm of fuzzy information processing, and does not import fuzzy logic computations into a neural network framework for advanced scattered data modelling, or function approximation.

The novel approach implies significant features, where both the transparency of fuzzy information processing and the ability of neural networks to handle complexity are uniquely combined. This was previously not established in this distinct form.

In fuzzy logic expert knowledge is involved. This is suitable in many circumstances, in particular in design, since the amount and quality of data available to form a model for fitness evaluation, are usually not enough to yield adequate models. The expert knowledge in fuzzy modelling is needed to decide about three main issues: (1) amount of membership functions to use (2) shape of the functions (3) and position of the functions in the universe of discourse. In the present approach of fuzzy neural tree, these three issues are determined in a different framework compared to the conventional Mamdani type of fuzzy modelling. Namely in place of deciding on a certain location for the membership functions by expert judgement, in the present case the location of membership functions is fixed, namely at unity of the input space of a node. Therefore the expert knowledge is brought into play in the present model not via the location of the membership function, but via the *weights* of the nodal connections. The weights are thus determined by an expert, and he/she is supposed to provide the weights. This way he expresses his/her knowledge about the relative importance among concepts that constitute the concept one level higher in the tree model. For a low number of sub-nodes such determination should be easy to make for an expert, and as the number of sub-nodes increases the method of AHP is an appropriate means to support the determination. Fixing the weights is thus equivalent to placing the membership functions. How many membership functions to place is also a decision by an expert based on his/her knowledge, namely how many sub-nodes are relevant to determine a node one level above. For each node in the tree the amount of inputs it has determines the amount of membership functions. There may be a large number of inputs considerable to associate with a certain node in general, however, many of these may be rather insignificant and therefore left out, since they will not influence the result in the fuzzy computations. An expert generally will decide the amount of inputs asking him-/herself the question: 'Which main aspects influence the performance of the design?', 'which sub-aspects influence these main aspects?' and so on, until he/she arrives at the level of physical properties of the design objects, which are subject to optimal identification. And it is emphasized that the designer should also specify how important the concepts within a sub-group are relative to each other. The shape of the membership functions is also considered. It is taken as a Gaussian to accomplish AND operation parsimoniously, and its width is determined by adaptive gradient optimization to ensure logical consistency in the model. This optimization procedure is a kind of *training* in the terminology of the neural-networks, however, in contrast to artificial neural networks its purpose is not to model knowledge inherent to data, but to ensure consistency within a given knowledge a designer/client has.

Thereafter the aim is to maximize certain outputs from that model simultaneously, in order to achieve a design with maximal performance. It is to be noted that the weights are to be fixed *prior* to such an optimization process, i.e. they are not subject to optimization together with the input parameters of the model. In the latter case the model would not be representing the knowledge at hand, but it becomes an arbitrary model. This means that outputs being optimized are not representing the relevant design features that should be optimized, making the effort invalid.

3.5. Conclusions

A novel fuzzy neural system as model for performance evaluation is described. The novelty is due to the particular neural tree structure bearing fuzzy logic concepts.

The fuzzy neural tree approach may be a significant step for design, because it models design criteria dealing with the common softness of requirements.

It is emphasized that the approach is unique in that it deals with the linguistic nature of the aspects involved, which are conventionally considered incomputable. This way available expert knowledge is made use of in its original and natural framework, as the natural imprecision of the human thinking is effectively mimicked. This is significant for performance evaluation in architectural design, since most design requirements include linguistic concepts, such as *functionality* or *perception aspects*.

While the model deals with the complexity of the criteria, transparency in the model is maintained at the same time. The transparency feature is significant, since it permits to insert expert knowledge into the model, while the fuzzy neural approach tolerates the characteristic imprecision of this knowledge. This way performance evaluation is accomplished using *human-like reasoning.*

Due to the local modelling properties of the Gaussian function used in the model, the complexity of the model can be increased without restriction. This means the model may absorb diverse aspects making comprehensive performance evaluation, involving criteria spanning various domains of expertise for instance. Using this approach an architect is able to assess the performance of his design using broad and deep knowledge at the same time. This is conventionally problematic due to the complexity of criteria, making consistent assessments difficult to achieve. The results provided by the model are deemed to be objective within a capacity proportional to the provided knowledge.

The essential added value of the novel method is summarized as enhanced ability to take knowledge from various parties into account resulting in more confidence on decisions through deep knowledge, higher precision, and greater transparency.

From efficiency viewpoint it is emphasized that in the novel neuro-fuzzy system no rules or data are needed to establish the model. Expert knowledge is inserted directly into the model. This is especially relevant for applications of the approach in cooperative design.

The results form the evaluation model described are to be used in a design process to arrive at designs satisfying the criteria most suitably. This is treated in the following chapter.

3.6. References

1. Ciftcioglu, Ö., Bittermann, M.S.: Solution diversity in multi-objective optimization: A study in Virtual Reality. World Congress on Computational Intelligence WCCI 2008, Hong Kong (2008)

2. Ciftcioglu, O., Bittermann, M.S., Sariyildiz, I.S.: Building performance analysis supported by GA. In: Tan, K.C., Xu, J.X. (eds.): 2007 IEEE Congress on Evolutionary Computation. IEEE, Singapore (2007) 489-495
3. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: A neural fuzzy system for soft computing. NAFIPS 2007. IEEE, San Diego, USA (2007) 489-495
4. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Fuzzy neural tree for knowledge driven design. The 2nd Int. Conf. Innov. Comp., Inf. and Contr. - ICICIC 2007. IEEE, Kumamoto, Japan (2007)
5. Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, Ö.: Performance-based Pareto optimal design. In: Rusák, I.H.a.Z. (ed.): TMCE 2008, Izmir, Turkey (2008)
6. Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, Ö.: Multi-objective optimization in the construction industry. AEC 2008, Antalya, Turkey (2008)
7. Bittermann, M.S., Ciftcioglu, O.: A cognitive system based on fuzzy information processing and multi-objective evolutionary algorithm. IEEE Conference on Evolutionary Computation - CEC 2009. IEEE, Trondheim, Norway (2009)
8. Augenbroe, G.: Integrated building performance evaluation in the early design stages. Building and Environment 27 (1992) 149-161
9. Mahdavi, A., Lam, K.P.: Performance simulation as front-end tool for integrative conceptual design evaluation. 2nd Int. Conf. of the Int. Building Performance Simulation Association (IBPSA), France (1991) 185-192
10. Beck, V.: Performance-based fire engineering design and its application in Australia. Fifth Int. Symp. of the International Association of Fire Safety Science, Melbourne, Australia (1997) 23-40
11. Hensen, J.L.M., Nakahara, N.: Building and environmental performance simulation: current state and future issues. Building and Environment 36 (2001) 671-672
12. Hartkopf, V., Loftness, V.: The concept of total building performance and building diagnostics. Automation in Construction 8 (1999) 377-393
13. Hien, W.N., Lam, K.P., Feriadi, H.: The use of performance-based simulation tools for building design and evaluation - a Singapore perspective. Building and Environment 35 (2000) 709-736
14. Zadeh, L.A.: Information granulation and its centrality in human and machine intelligence. Rough Sets and Current Trends in Computing, Vol. 1424/1998. Springer, Heidelberg (1998) 35-36
15. Kanda, J., Shah, H.: Engineering role in failure cost evaluation for buildings. Structural Safety 19 (1997) 79-90
16. Castilho, V.C.d., Debs, M.K.E., Botelho, C.: Using a modified genetic algorithm to minimize the production costs for slabs of precast prestressed concrete joists. Engineering Applications of Artificial Intelligence 20 (2007) 519-530
17. Monks, M., Oh, B.M., Dorsey, J.: Audioptimization: goal-based acoustic design. IEEE Computer Graphics and Applications 20 (2000) 76-90
18. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological review 63 (1956) 81-97
19. Baddeley, A.D.: Working Memory. Oxford University Press, New York (1986)
20. Saaty, T.L.: The Brain: Unraveling the mystery of how it works,. RWS Publications, Pittsburgh (2000)
21. Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)
22. Ciftcioglu, Ö., Sariyildiz, I.S.: The analytical hierarchy process applied for design analysis. 3rd International Conference on Innovation in Architecture, Engineering and Construction - AEC 2005, Rotterdam, The Netherlands (2005)
23. Zangemeister, C.: Nutzwertanalyse in der Systemtechnik. Wittmann'sche Buchhandlung, Munchen (1971)
24. Zadeh, L.A.: Fuzzy logic, neural networks and soft computing. Communications of the ACM 37 (1994) 77-84
25. Hunt, K.J., Haas, R., Murray-Smith, R.: Extending the functional equivalence of radial basis function networks and fuzzy inference systems. IEEE Trans. Neural Networks 7 (1996)

26. Kasabov, N.K.: Foundadtions of Neural Networks, Fuzzy Systems, and Knowledge Engineering. MIT Press, Cambridge (1996)
27. Broomhead, D.S., Lowe, D.: Multivariable functional interpolation and adaptive networks. Complex Systems 2 2 (1988) 321-355
28. Zadeh, L.A.: Fuzzy logic and approximate reasoning. Synthese 30 (1975) 407-428
29. Zadeh, L.A.: Fuzzy sets. Information and Control 8 (3) (1965) 338-353
30. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. Systems, Man and Cybernetics 3 (1973) 28-44
31. Mamdani, E.H.: Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Trans. Computers 26 (1977) 1182-1191
32. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics 15 (1985) 116-132
33. Bellmann: Adaptive control processes: A guided tour. Princeton University Press, Princeton (1961)
34. Jang, J.S.R., Sun, C.T.: Functional equivalence between radial basis function networks and fuzzy inference systems. IEEE Trans. Neural Networks 4 (1993) 156-158
35. Foresti, G.L., Micheloni, C.: Generalized neural trees for pattern classification. Neural Networks, IEEE Transactions on 13 (2002) 1540-1547
36. Sankar, A., Mammone, R.J.: Neural tree networks. In: Mammone, R.J., Zeevi, Y. (eds.): Neural Networks:Theory and Applications. Academic, New York (1991) 281 -302
37. Sirat, J.A., Nadal, J.P.: Neural tree networks. Network 1 (1990) 423-438
38. d'Alché-Buc, F., Zwierski, D., Nadal, J.P.: Trio learning: a new strategy for building hybrid neural trees. Int. J Syst. 5 (1994) 259-274
39. Schacter, D.L.: Memory. In: Posner, M. (ed.): Foundations of Cognitive Science. The MIT Press, Cambridge, MA (1989)
40. Lawson, B.: How Designers Think. Butterworth-Heinemann, Oxford (1997)
41. Hamel, R.: Over het Denken van de Architect. AHA books, Amsterdam (1990)

4. Design generation

In the previous chapter a model to evaluate the performance of designs is presented. This chapter describes how the evaluation model is used in the process of generating designs that satisfy the design criteria. The chapter is based on the publications [1-5].

Section 4.1 describes the challenges in the process to generate suitable designs and the shortcomings of existing computational approaches. In section 4.2 the favourable properties of evolutionary computation as methodology for design generation are described. Architectural design involves the satisfaction of several criteria simultaneously. Therefore the concept of Pareto optimality plays an important role in design generation. The Pareto concept is explained in section 4.3. A state-of-the-art challenge concerning Pareto-based optimization methods is the maintenance of solution diversity. This plays a significant role in architectural design, because it is desirable to obtain a number of alternative solutions with different characteristics, yet comparable performance. This is in order to gain awareness of the trade-offs inherent to the task at hand. The issue of diversity preservation is addressed in section 4.4. A novel approach is presented that enhances the Pareto ranking of genetic algorithms.

In section 4.5 the intelligent design objects approach is presented, where the three soft computing components described in the previous chapters are combined. The components form a novel computational design system that exhibits some cognitive behaviour. The interaction between designer and the system is described forming a cognitive approach for performance-based architectural design. This concludes the theoretical part of the thesis.

4.1. Challenges for design generation

4.1.1 Challenges due to multiple objectives

When we are designing we need to verify the performance of our design with respect to our criteria. This means we consider various performance aspects of our design. We investigate its functional performance, technical performance, energy performance, aesthetic performance, etc. We do this in order to know how good our design is overall, and where there is potential for increasing its performance.

Judging the overall performance of a design is challenging. This is because the criteria involve a large amount of relations between performance attributes and the variables of a design. For example a design's performance is influenced by a number of performance aspects, such as its functional, technical, perception related performance. Each of these aspects depends on the satisfaction of a number of elemental requirements. For example functionality is influenced by the satisfaction of the requirements for sufficient size of the spaces, requirements for high accessibility etc. Satisfaction of these sub-aspects is again depending on multiple design variables.

Another complication is that the same design variable may influence different performance aspects at the same time. For instance when we change the position of a window, this influences both the visual perception properties from the exterior, from the interior, and it may also influence the energy use of the building. Some of the requirements imposed on an object are conflicting. For example a wall may be expected to provide visual separation, climate insulation, noise absorbance, low manufacturing and material cost, as well as high reusability simultaneously. Clearly some of the requirements are conflicting, for instance low material cost and high insulation.

In the previous chapter a method was presented for performance assessment. The method resolves a bottleneck that prevented comprehensive evaluation of performance in architectural design up till now. This bottleneck concerns the treatment of requirements that are commonly satisfied to a degree, and not merely fully or not. This is a common property of design requirements. The reason is that some requirements are conflicting, so that we cannot have all requirements maximally fulfilled. The method presented in the previous chapter is able to handle a multitude of such soft design requirements at the same time. This is a formidable task when conventional means are used.

Once the model is established, providing the features of a design at the input of the model, at the model outputs the design's performance score is obtained.

4.1.2 Challenges due to excessive solution options

The information obtained from the performance evaluation is useful to search for designs having superior performance. However this process is not straight forward. Assuming we know that the performance of our design is moderate overall; it remains a difficult question how to increase its performance. Conventionally we play with the parameters of the design, for example change the overall shape of the building, change its orientation, modify the spatial organization etc. Then for every modification we observe the change of the performance of the design, until we are satisfied.

However, a design consists of many design objects, such as several spaces, walls, floors, ceilings, etc., and each object may have several parameters characterizing the object. For example a room is characterized by its size, proportion, orientation, and its material properties. There are many possible options to decide on even regarding such a basic design object. Next to this, in the last decades the technological means to realize a building multiplied. New building materials and building products emerged, such as composites, and manufacturing and erection technology advanced. This means the designer has an excessive amount of options to consider in combination. Each of the options has their unique influence with respect to the satisfaction of the design requirements.

As it is difficult to foresee which combinations will have the most positive effect on the design performance, we cannot afford to evaluate only a small portion of the alternatives. The amount of possible combinations of parameter values we should consider to be confident not to have missed a potent option is very large. However, due to time limitations conventionally we are able to investigate only a limited amount of options. It is clear that this bears the risk that viable alternatives remain unexplored. Another issue is that when we are exploring the solution alternatives we should be aware of the dependencies among the design objects. This means we should remember all the relevant criteria, the dependencies among the design variables and the performance aspects to be sure we are moving in the right direction with our design actions. It is clear that it is problematic for us to bear many relations in mind and at the same time think of improved alternatives.

Therefore it is difficult to foresee, which design modification will bring about the greatest increase in the overall performance. This is problematic because a certain design modification may apparently improve the design, because it increases the design's performance on a certain aspect; however the same modification may have a negative effect with respect to other aspects. For example a slight change in the orientation of a building may have a positive effect on the perceptual properties of the interior; however it may have a strong negative impact on the energy consumption of the building. Or changing the position of a column may improve the structural performance greatly, while it may have a strong negative effect on functionality or perceptual aspects at the interior. The complexity of design requirements and the large amount of possible solutions makes it uncertain that we find suitable solutions that satisfy our criteria as we desire.

Due to this uncertainty designers are forced to some extent to follow ‘trodden paths’ to arrive at solutions. That is, traditionally designers deal with the large amount of solution options and complexity of criteria by segmenting the problem into parts. For example first we deal with the general shape of the building and afterwards we deciding about the details [6, 7]. However it is well known that when we fix a certain property of a design at a certain moment during the process without taking the full range of relevant aspects into account, afterwards this decision may turn out to be inferior. This is because the details of a design have an influence on the performance of the whole and vice versa. Due to the large amount of dependencies among design objects, implications of the first decision may not be fully overseen. Therefore aspects become apparent that have an influence on the previous consideration. It may turn out our previous decision was not viable. As a result, a design is obtained that does not match the intended purpose to the desired extent. In late phases of a design process problems occur that cannot be resolved easily anymore due to premature previous commitment [8]. This problem is indicated by the large amount of failure costs the building industry produces due to planning errors. That is, certain aspects were not taken into account duly enough, or thoroughly enough during design, to ensure delivery of a flawless building.

Due to this uncertainty it is problematic to ensure that the solutions we generate match the given criteria. This issue is an essential motivation to bring computation into the play. However, the complexity of design makes it challenging for computational methods to accomplish this task.

4.1.3 Computational issues to deal with the challenges

Due to the uncertainty about the consequences of a design modification, generally rule-based approaches for design generation have limited effectiveness. To exemplify this statement let us consider the following simple application of a rule. Say we would prefer a space to be a certain size, for example 60m^2, and if the space is less than this size we would increase it until we reach the required amount. Such a rule is problematic in case that the size of the room influences other factors of the design as well. For example, this size may make the building too expensive or it may make the functional organization problematic. Such multiple effects of a design modification generally occur in design. Therefore, if we blindly apply a given rule to increase the performance of a design, we may not be sure that this will result in a more suitable design. We generally have to consider the entire situation and aim to modify many parameters in parallel to understand what the most viable direction to follow is. When we are looking for a design with a high performance it is clearly not enough to apply simple rules. However, due to the large amount of possible combinations of parameter values that we ought to investigate, we need to consider methods to accomplish this that can warrant success to some extent.

In computational terminology a design solution is a set of objects with a specific pattern of values assigned to the objects’ parameters within their respective boundaries. The boundaries define a region in the solution space, beyond which exploration of alternative solutions is deemed to be irrelevant. For example, if it is known that a house should be positioned within a given lot, it is irrelevant to consider positions beyond the lot boundaries. Every possible pattern, i.e. every combination of different parameter values within the boundaries, is another possible design alternative. The essential purpose of design is to find those designs that satisfy the criteria to the greatest extent. Ultimately we aim to create designs with maximal design performance.

As there are many parameters in a design, and each one may take several possible values, it is clear that the amount of possible solutions is enormous. This is known as *combinatorial explosion*. The amount is usually so large that it is not feasible to investigate every possible solution, even if only a very short time is spent for the evaluation of a solution's performance. This is valid even in the era of fast computers. The reason is that already with few design parameters the result of the computation would not be available in time. This is clear from the following consideration.

Say a design has 20 parameters and each one has 10 possible states, then the amount of possible, unique combinations is 10^{20}. Assuming that it takes a computer merely a single processing cycle to investigate the performance of one of the possible combinations, and the processor has a frequency of 10 GHz it would take 10^{10} seconds to consider the solutions, which is more than 300 years.

4.1.4 Search for maximal design performance

Let us consider we established a model to evaluate the performance of design solutions as described in chapter 3. Then we wish to arrive at those designs that have the highest performance score according to our model. With this understanding design is an optimization problem. In particular the problem is to maximize the design performance by identifying the suitable values for the design variables.

Optimization is a procedure of finding and comparing feasible solutions until no better solution can be found [9]. It is an important area of scientific interest, and there are numerous optimization methodologies that are suitable for different optimization problems. For architectural design a number of typical characteristics of our task favour a certain optimization methodology over the others. These issues concern the complexity of the design criteria and the combinatorial explosion in the solution space. The complexity refers particularly to the large amount of interdependencies among design aspects and the conflicts among requirements. Due to the complexity, when we modify a design object, for example we change the location of the entrance to the building; the performance may increase for a certain location. And as we continue to move the entrance the performance may decrease

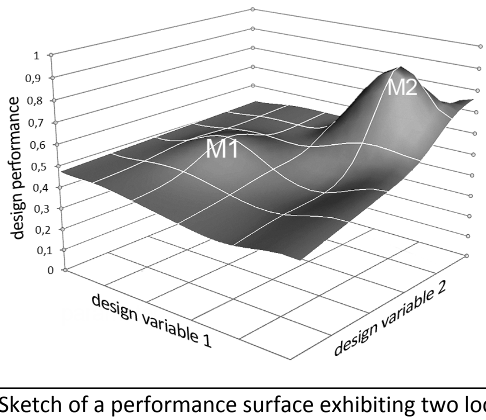


Fig. 4.1 Sketch of a performance surface exhibiting two local maxima

again. And as we continue to displace the entrance it may increase again for another location, etc. This means the relation is non-linear in general. It is noted that the specific resulting performance score is depended on the present state of the other design variables. For example the same modification of the entrance will yield a different effect on the design performance, when the building layout is different than before. It is emphasized that the relation between design performance and design variables is non-linear in general, and it may be highly non-linear when the complexity of our design task increases. This non-linearity is illustrated in figure 4.1, where the fictitious relation among two design variables and a corresponding design performance is plotted. The resulting surface represents the performance for every possible combination of values for design variable 1 and 2.

Generally a performance surface can exhibit local maxima or minima. A local maximum of the performance surface refers to a certain location in the parameter space, i.e. a certain design, where the corresponding performance is relatively higher compared to regions around the location. This is exemplified in figure 4.1 by means of the two humps labelled *M1* and *M2*, where each one represents a local maximum of the performance.

To illustrate the meaning of such local maxima let us consider for example that design variable 1 and 2 refer to the *x* and *y* coordinates of a column respectively. Considering the figure, apparently there are two outstanding locations *M1* and *M2* for the column, at which the design requirements are satisfied more than for other possible locations. Comparing the outstanding locations, clearly location *M2* is preferred over *M1*, as it yields a higher performance value.

A challenge for optimization methods is avoiding to be trapped at a local maximum, such as *M1*, while missing that there is an even better place *M2*, where the design performance is higher than at *M1.* Let us assume the surface in figure 4.1 represents the entire search space, then *M2* is termed global maximum. This means the design at M2 is the prominent solution. Successful optimization has to be able to evade local, inferior places, such as *M1*, and identify the solutions with the maximal design performance in a global sense.

Optimization methods that are based on direct calculus, such as hill-climbing algorithms are unable to handle such a situation. This is because they are inherently local. Clearly

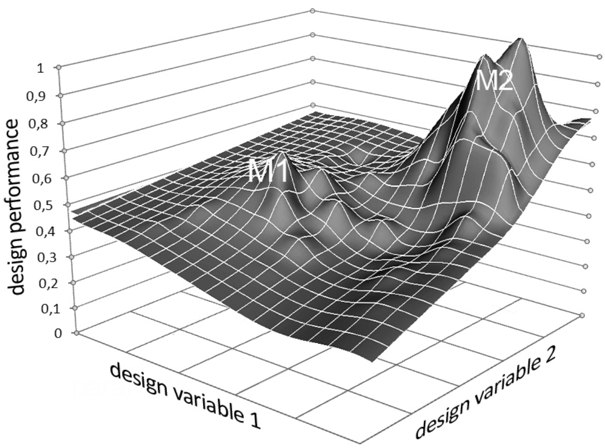


Fig. 4.2 Sketch of a performance surface with multiple local maxima

starting the search at the neighbourhood of a lower peak, will cause the search to miss the main peak. Once the lower peak is reached, further improvement is only obtained through random restart or other ad-hoc endeavours [10].

Figure 4.2 shows a second fictitious design task. In this example the performance surface has many local maxima. This is referred to as *multimodal* fitness.
We can characterize this topology of the objective space as *rougher* compared to figure 4.1. In such a rough fitness landscape it is clearly more difficult for an optimization method to avoid being trapped near a local maximum. This is because there are many local maxima, and they occupy relatively little space with respect to the design variables.

A problem with many conflicting requirements has a multimodal fitness landscape. This is usually the case in design. This means in design we generally expect to have performance models with highly non-linear relations among design variables and a corresponding performance score. In other words we can expect design problems to resemble more to the surface in figure 4.2 rather than figure 4.1.

Therefore design requires computational methods that are equipped to handle multimodal fitness landscapes. This entails that calculus based methods are generally unsuitable for design generation. As remaining option one may consider enumerative schemes or random walks for finding designs with maximal performance. Enumerative schemes are based on evaluating a number of possible solutions one by one. Random walks are based on moving through the space of solutions randomly. The efficiency of these methods drastically decreases with increasing size of the search space. This is referred to as the curse of dimensionality [11]. As search spaces in design are generally large, both methods are unsuitable for design.

As a bottom line we can conclude that calculus based methods, as well as enumerative schemes and random walk algorithms are generally unsuitable for architectural design.

4.2. Evolutionary search

In the last decades a search methodology emerged in the domain of computational intelligence, which is able to handle the complexity issues of design generation. This means the methodology is able to deal with large parameter spaces, discrete parameters, and evades local maxima on the performance surface. Due to these features the methodology is outstandingly suitable for design generation. The methodology is known as *evolutionary search* [10].

4.2.1 Concept of evolutionary search

Evolutionary search is based on exploring the space of possible solutions starting with a collection of random solutions. Therefore the methodology belongs to the class of *stochastic* or *randomized* search methods. It is clear that by means of considering only a sample of the possible solutions the methodology sacrifices completeness of the information for the ability of concluding in time. This way it deals with the limited computational resources available and still is able to explore a vast search space. It is emphasized that randomness is the methodology's essential asset in dealing with the vastness of the search space. Next to this, additional features enable the methodology to deal with the problems design tasks pose to search methods, which impede the traditional calculus-based and enumerative approaches.

The methodology is inspired by mechanisms in biological evolution that enable the survival of a species in a complex environment [10]. This is how the methodology obtained its name. Essential abilities of biological species to survive are their ability to encourage well performing parents to mate. Through combination of their genetic material offspring of such individuals is expected to have a higher chance of performing well in the environment, too. Hopefully it even outperforms its parents. This principle is used in evolutionary computation. Specifically the evolutionary method known as *genetic algorithm* is considered, as it is able to deal with the complexity issues mentioned above. In particular it is suitable to handle the multiple objectives inherent in design.

A genetic algorithm (GA) stores a design solution in the form of a *chromosome* that consists of *genes*. A number of solutions form the set of possible solutions of the optimization. This set is termed a *population*. Each gene represents a value that belongs to a specific design variable. For example a genetic description of a room may contain the size of the room, the position of the windows, doors, materials of the walls etc. This information is stored in one long chain of values termed *chromosome*. Usually in GA the values are represented in binary form, i.e. as a string of bits.

4.2.2 Design generation using genetic algorithm

An initial population of designs generated by a genetic algorithm is a random combination of the variables of the design within their respective boundaries. The chromosomes in a population are then graded using a criterion termed *fitness*. The fitness quantifies the performance of a solution.

The fitness evaluation marks the end of an iteration of the algorithm. According to the evaluation results a new population is formed. Proportional to its fitness, each chromosome is given a chance to survive to the next generation. This procedure is termed *reproduction,* in analogy to biological evolution. In a natural way, in the consecutive population there are more individuals inclined to be acceptable solutions than the ones in the preceding population.

The ensuing population is further prepared using genetic operators. These operators are *crossover*, and *mutation*. The operators are exemplified in figure 4.3. The genetic operators are applied on the parents, which are probabilistically selected based on their fitness. This way a population of new possible solutions is generated called *offspring*. The objective of the crossover operation is to explore different places in the search space, while maintaining some of the features of the parent solution found earlier. The mutation operation injects some randomness into the population to avoid that the variation among chromosomes is reduced too quickly. This is to prevent that the search becomes trapped at an optimum that is local and inferior compared to the global one.

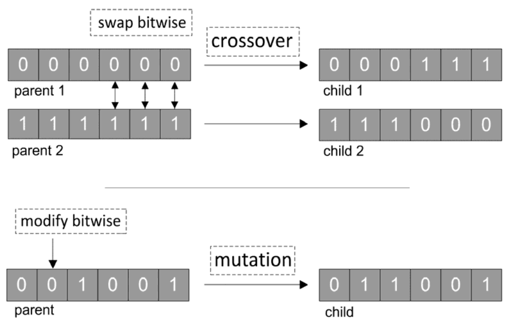


Fig. 4.3 Genetic operators

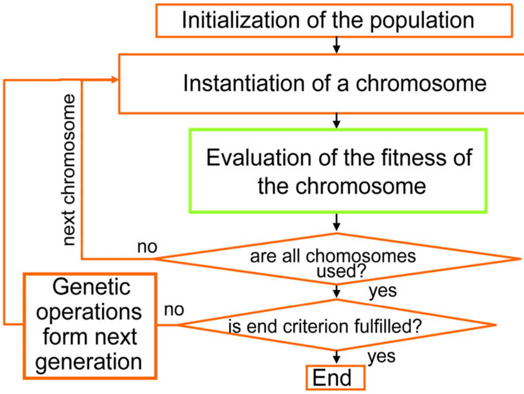


Fig. 4.4 Flowchart of a genetic algorithm

The implementation of a GA is as follows. We assume that, the initial population is created randomly at time *t=0*. Let *P(t)* be the population at time *t*, consisting of an amount of chromosomes *P* as a constant scalar quantity. The population size is to be selected in accordance to the complexity of the problem at hand, as it influences the performance of the algorithm.

The crossover operation is executed with an associated probability P_C recombining two chromosomes. This is dome by cutting them at a random position and partially exchanging the genes of the chromosomes. It is significant that the value for P_c is not set too high, in order to avoid premature *convergence* of the search. The value should not be too low either, to ensure that the search is converging. Convergence refers to the event that the average fitness of the population is relatively close to the performance of the best chromosome that was found during the search. It is noted that the convergence of the search is an essential performance measure for a genetic search.

Mutation, with an associated probability P_M changes the values of some randomly selected genes. The value for P_M should not be set too high, to avoid erratic behaviour of the search. It should neither be too low to avoid premature convergence.

Then the *fitness* of each newly formed chromosome is evaluated. Chromosomes with a low fitness score have a high chance of being replaced with those having high scores. This is the *reproduction* process. By the end of the reproduction a new population is formed, namely population *P(t+1)* at the time *t+1*. We note that the reproduction ensures that the population size is maintained as constant. It is emphasized that the application of GA requires careful adjustment of the genetic parameters. These parameters are population size, crossover probability and mutation probability. Each parameter greatly influences the performance of the GA. More details on GA can be obtained from the literature [10].

In order to enable a designer to visually monitor the search process, for instance to verify the appropriateness of the search boundaries, an additional step is integrated into the above algorithm. This step is *instantiation* of the search process. This means the features that are generated by the GA are assigned to the design objects continuously during the search. The integration of the instantiation procedure into a GA is shown in figure 4.5. In particular the instantiation is executed for the solution with the highest performance the search finds. This way the objects assume the features of the best solution, while the non-evolved object features are also instantiated. As result the design

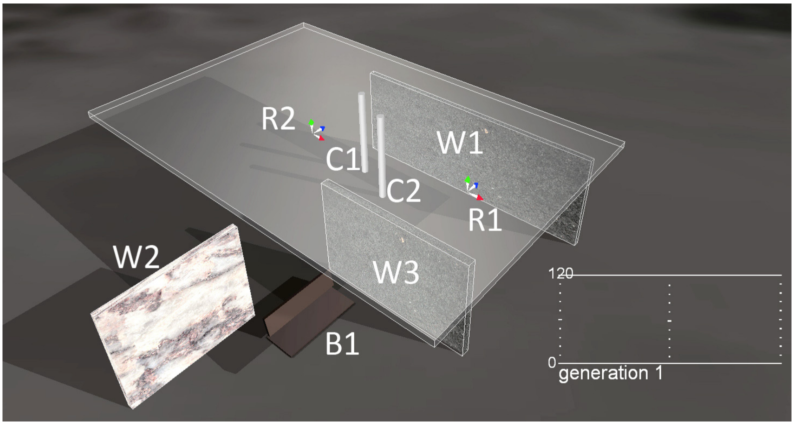


Fig. 4.5 An initial design from the first generation of the GA

objects exhibit goal-oriented behaviour, i.e. they can be observed as they gradually improve their fitness.

As an example, let us consider that the position of a wall is subject to genetic search. Instantiation of the wall means that it is represented at the evolved position including its other, non-evolved, features, such as length, height, width, and material.

This feature has some merits allowing designers to monitor the search, for example to verify if the search boundaries put are well chose. Beyond that this novel feature is imperative when the ray-trace perception analyses described in Chapter 2 is used in the performance assessment. In such a case object inherent features that may not be subject to modification by GA, such as width length of an object, may play an essential role in fitness assessment. In other words, in order to make the perceptual assessments the environmental geometry should be present to execute the measurements. In this case all chromosomes of a population are subject to instantiation during the fitness assessment procedure.

4.2.3 Example of a genetic search process

To exemplify the implementation of genetic algorithms let us consider the following demonstrative exercise.

The exercise concerns the creation of a simple pavilion that consists of a number of design objects. The objects are three walls, two columns, a couch and a roof slab, seen from figure 4.5. Two reference locations *R1* and *R2* are indicated in the figure as well. These are the centres of two spaces to be enclosed by the walls. Please note that the reference objects *R1* and *R2* are fixed locations in the environment.

The task for the genetic algorithm in this exercise is to place the objects to positions, so that the distances among the objects become desired values. The distances are measured between the centres of gravity of the objects for simplicity. The desired distances are specified in table 4.1. The distances were selected, so that there is at least one solution that satisfies the requirements. From the table it is seen that for example the bench *B1* is required to be very near to wall *W2*, namely 0.8m, while at the same time it should be at 2.5m distance from *R1*. The goal in the present task is to minimize the difference between the desired distances and the actual distances among the objects. The difference is also referred to as *error*, and it is essential in the fitness assessment of the GA: the reciprocal of the error is used as fitness value of the chromosome, so that the fitness function is given by

Table 4.1 Desired distances among the design objects, the values are in m

Object	W1	W2	W3	C1	C2	B1	R1	R2
W1		2.8	4.5	5.6	9.7		3.4	3.7
W2	2.8		3.3	3.3	6.8	0.8	3.4	2.3
W3	4.5	3.3		5.9	4.7		2.2	5.3
C1	5.6	3.3	5.9		9.7			2.1
C2	9.7	6.8	4.7	9.7			3.6	
B1		0.8					2.5	

$$F(\boldsymbol{d}) = \left[\sum_{i=1}^{i=20} \left| \left(d_i - d^*_i \right) \right| \right]^{-1} \quad (4.1)$$

where d_i denotes a certain distance between two objects in the scene for the *i*-th requirement; d^*_i denotes the corresponding desired distance.

Figure 4.6 shows how the GA minimizes the error during its search. From the figure we note that after about 80 generations the error of the best solution found by the GA and the average error approach each other. This is the *convergence* described above, and it is a measure of the performance of the algorithm.

Figure 4.7 shows the behaviour of the design objects aiming to satisfy the requirements. The figure shows the best solution that was found during the search. Figure 4.7a shows the design in the initial population, figure 4.7b after 8 generations and 4.7c after 20 generations.

Considering figure 4.7 we note that the random configuration in the initial generation of the search gradually develops into the final design, shown in figure 4.8. The design in figure 4.8 is found after 90 generations. It has a total difference to the demanded distances of about 2.4m. With 20 distance requirements to be fulfilled the average error is about 12cm per requirement. This result can still improve, however, the improvements becomes less and less over time. By monitoring the fitness a designer is informed about the state of the search process in real-time.

It should be noted that the fitness evaluation of the present example needs further consideration. The design requirements used in the example are unusual in the sense that they exclusively consist of requirements for certain values of object parameters. In design

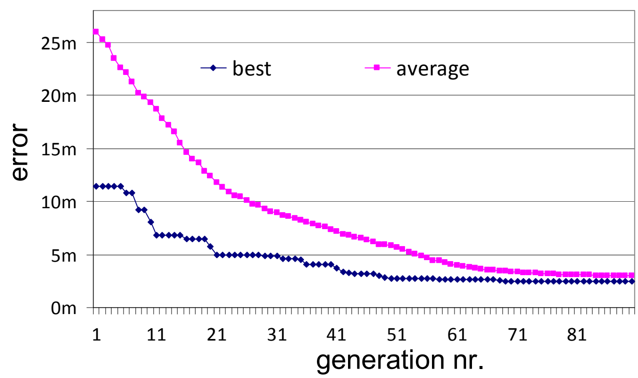
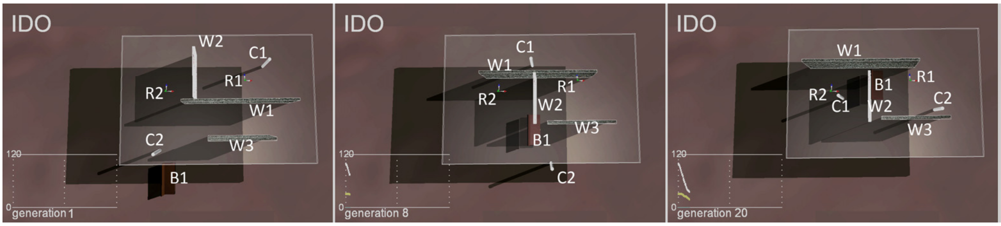


Fig. 4.6 Genetic search process



(a) (b) (c)

Fig. 4.7 Goal-oriented object behaviour due to instantiation of the genetic search process

generally requirements are expressed in terms of abstract qualities, such as functionality or spaciousness. The latter qualities are complex functions of the objects' parameters, and there is generally some tolerance for their partial satisfaction associated with them.

Another issue is that in the demonstrative example the requirements are treated as equally important. Usually in a design there are differences among requirements in terms of their relative importance for the fitness of a design. In order to express these differences one may consider introducing a weight factor for each requirement, so that Eq. (4.1) becomes

$$F(\boldsymbol{d}) = \left[\sum_{i=1}^{i=20} \left| \left(d_i - d^*{}_i \right) \right| w_i \right]^{-1}, \tag{4.2}$$

where w_i expresses the importance of the i-th requirement. However, with a large number of requirements it becomes formidably difficult to specify accurately the relative importance among requirements with respect to the total performance of a solution.

4.3. Pareto optimality

4.3.1 Single objective vs. multi-objective search

When we start to design we have many criteria in mind. Our building should be functional, aesthetically pleasing, energy efficient, cost efficient etc. These criteria are in conflict with each other in general. For example a low energy design is generally also more expensive than. At the outset of design we are usually not certain of the relative importance among the criteria. A most common question we ask ourselves during design is how much attention we shall put to functionality and how much to aesthetics. The underlying question is how important the criteria are relative to each other. The reason why this is a problematic question to answer is because it is vague what we exactly mean by the criteria.

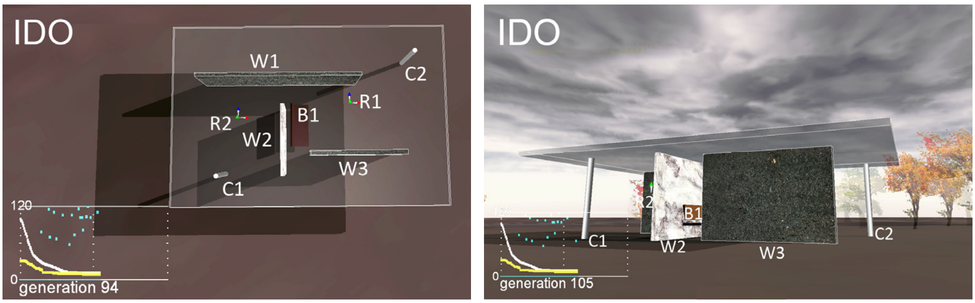


Fig. 4.8 Resulting optimal design

Functionality for example is a linguistic concept that refers other concepts and eventually via those it refers to physical properties of design objects like sizes, widths etc. Due to the complexity of the meaning of the concepts forming our design criteria it is difficult to grasp this meaning when we aim to compare their significance in a project. As we cannot pin-point relative importance among goals in general in design, this means we are essentially not in a position to execute an optimization with the single goal, being maximal overall performance. We have to use other means to reach our ultimate goal.

In many design applications optimization it is treated in single objective form [12-19]. In these cases the objective is known as minimization of the *cost function,* which is equivalent to maximization of the *fitness function.* In such approaches multiple objectives are combined in one function. This way they form a single objective to be optimized. This is also the case in the example in the previous section, where the single objective is to maximize a fitness function. This approach is suitable when we are able to specify the relative importance among the design criteria with certainty. However we may not be able to do that, due to the complexity of the criteria as mentioned before. For instance when we consider how important the functionality of the building is with respect to its sustainability we are not easily to specify this with certainty. Both aspects are important and it is not easy to express the relative importance before knowing the exact implications this has on corresponding solutions. This is particularly problematic in design, since the amount of requirements and their imprecision is so high that the relative importance among requirements may not be assessed with sufficient accuracy before having more information on the tradeoffs inherent in the design task. In single-objective optimization it is necessary to commit to a single fitness function prior to knowing the tradeoffs among potential solutions. This is a significant drawback of this approach for design.

Therefore, as an extension of single objective optimization an optimization approach was established, where one or more objectives are simultaneously satisfied, next to the minimization of the cost function [9]. In this approach the functions involved are *simultaneously* minimized (or maximized) and a number of applications of the approach in architectural design can be found in the literature [20-25]. The information on relative importance among criteria is not necessary. This is termed *multi-objective*, or *vector-valued* optimization, and single-objective optimization is considered a sub-problem of *multi-objective* optimization [9].

The essential difference between single-objective optimization and multi-objective optimization is that there is no single optimal solution in the latter case. In the multi-objective case there are multiple solutions, which are in some sense equally valid. Using a single-objective optimization approach, in one run of the algorithm merely one solution is obtained. This means the result from the endeavour is not very informative regarding the validation of criteria. Obtaining multiple solutions of a multi-objective optimization problem in one run of an algorithm is a significant desirable feature for design, and this is uniquely accomplished by means of evolutionary computation.

Based on the multiplicity of solution alternatives the decision making remains flexible. That is, a designer is able to postpone his/her commitment regarding the relative importance among the design criteria. This is postponed until he/she is more aware of the trade-off characteristics inherent to the design.

4.3.2 The concept of dominance among solutions

To deal with multi-objectivity, evolutionary algorithms with genetic operators are effective in defining the search direction for rapid and effective convergence. Basically, in a multi-objective case the search direction is not one but there may be many directions. Therefore, during the search a single preferred direction cannot be identified and even this is not desirable. In the evolutionary computation case a population of candidate solutions can easily hint about the desired directions of the search. This way the candidate solutions appearing during the search process become more probable to satisfy the ultimate goal.

For example, let us say we wish to maximize the comfort of a building and at the same time ensure the building is inexpensive. Initially we are uncertain how important to weigh each of these aspects. This is because we are unaware how much we have to pay for a gain in comfort. Therefore we look for a number of solutions that represent all optimal compromises among the two criteria. One solution should for example be very inexpensive while the corresponding comfort should still be as high as we can achieve for the low price. On the other extreme we should look for an extremely comfortable building, while it should be as inexpensive as we can achieve at same time. Clearly we should not be satisfied with a solution that is also extremely functional as the one mentioned before, but which could be still much less expensive. Such a solution we term inferior with respect to the previous one. Clearly it is not reasonable to include this solution in our final consideration, provided the two criteria are all we are concerned about. In between the two extreme solutions, there is a whole range of solutions that are all interesting compromises in the sense that no other solution outperforms them in both respects at the same time. Among these solutions we are likely to find a solution suiting our ultimate preferences. Establishing these solutions is explained as follows.

With the advent of evolutionary algorithms in the last decades, multi-objective evolutionary algorithms (MOEAs) [9, 26-29] are extensively being investigated for solving multi-objective optimization problems. Evolutionary algorithms are particularly suitable for this, since they evolve simultaneously a population of potential solutions.

A particularly desirable feature of solutions pursued in multi-objective optimization is that they are *non-dominated*. A solution is non-dominated if there exists no other solution that performs better in every respect. In other words, when compared to any other solution, a non-dominated solution is superior for at least one criterion. Such a solution is also termed a *Pareto optimal* solution. As design performance is subject to *maximization*, let us consider an optimization problem where the objective function values are requested to be higher in the Pareto front, relative to those in the dominated region. That is, for any two solutions $\boldsymbol{d_1}$ and $\boldsymbol{d_2}$ of a design problem, $\boldsymbol{d_2}$ dominates $\boldsymbol{d_1}$, if

$$\forall i : f_i(d_2) \geq f_i(d_1) \wedge \exists i : f_i(d_2) > f_i(d_1),$$

where f_i is the *i*-th objective.

The set of Pareto optimal solutions of a multi-objective optimization problem form a surface in objective space. This surface is known as *Pareto surface* or *Pareto optimal front*. Establishing this front is the primary goal in multi-objective search. MOEAs are able to identify the Pareto front in a single run of the algorithm. This is an appealing property, and there is a fast growing interest in the method in the last decade.

4.3.3 Diversity of solutions

The Pareto front is a very important concept for architectural design, because a designer can delay his commitment for the relative importance among design criteria and make the commitment with great awareness. Although the concept of Pareto front is well known and significant, the formation of the Pareto front is not straightforward. This is because the strict search of non-dominated regions in the multi-objective solution space prematurely excludes some of the potential solutions. The reason for this lies in the Pareto ranking technique conventionally applied in MOEAs. This technique favours only a few solutions of the population, while it considers a large number of solutions to be unfavourable. Therefore, the diversity of the available solutions that underlie the formation of the next generation quickly diminishes resulting in aggregated solutions.

As consequence of the premature exclusion of potential solutions the remaining solutions aggregate in the solution space. Aggregation occurs, since the search does not maintain enough solutions that lie in between dominant solutions in objective space. Solutions with many objectives are usually non-dominated with respect to each other. This is clear considering that for many objectives it is an unlikely event that a solution is outperformed by another one in every respect. This means there is very low selection pressure toward the Pareto front in Pareto dominance-based evolutionary multi-objective (EMO) algorithms with many objectives [30]. This means a Pareto surface is not fully developed, and the diversity of the solutions on the front is not fully exercised.

Aggregation of solutions on the front is undesirable, because then an architect has an incomplete overview of the favourable solution alternatives. This means there may be desirable solutions hidden in the unexplored areas of the solution space, and a DM has less information to validate the relative importance among criteria or to exercise secondary criteria he/she has in mind. Therefore solution diversity is particularly desirable in architectural design.

To maintain diversity on the Pareto front often *sharing* [31] or *niching* [32] are used. These techniques aim at spreading out a population by decreasing the fitness of an individual solution depending on the closeness of its neighbours. In this way solutions in sparsely covered regions of the objective space are favoured over solutions in crowded areas.

It should be noted that sharing is a forced manner to obtain solution diversity. It requires the specification of a sharing parameter σ_{share} that greatly influences the effectiveness of the search. Although there is work on dynamic sizing of the sharing parameter [33], a diversity-preservation mechanism without such a parameter is desirable. In other words, sharing or niching aim to prevent aggregation *symptoms* on the Pareto front, while they do not address the source of the aggregation phenomenon. It is desirable to investigate more natural ways to let the search arrive at a non-aggregated Pareto front. This is a state-of-the-art issue in evolutionary optimization [34].

In the following, this issue is investigated and a solution is provided, so that diversity of multi-objective solutions is maintained on the Pareto front without sharing or niching [35]. The method used is termed *relaxed dominance* and it basically refers to a *degree* of dominance in the terminology of MOEAs.

4.4. A novel approach for diversity of solutions

4.4.1 Multi-objective optimization with a relaxed-dominance concept

Diversity of solutions is significant, so that a designer can decide about the relative importance among his/her criteria with great awareness. That is, the designer has already information how his decision will influence the solution. Some important features of the latest generation MOEAs address the selection of the potential solutions during the optimization process, and diversity-preserving strategies in objective space.

Next to the principles of GA optimization, in MO algorithms, in many cases the use of Pareto ranking is a fundamental selection method. Its effectiveness is demonstrated for a moderate number of objectives, which are subject to optimization simultaneously [36].

On the Pareto surface the solutions are different, but they are assumed to be equivalently valid in Pareto sense. This means they are all non-dominated. Selection of one of the solutions among those many is based on some higher order preferences. Such a higher order references means specification of the relative importance among the optimization criteria.

For conflicting criteria, which are commonly present in design tasks, increased satisfaction of one criterion implies loss with respect to satisfaction of another criterion. Such local gain at the cost of a loss elsewhere is termed *trade-off*. Decision makers usually are able to intuitively express the proportionality of the expected gain with respect to the loss [37]. This information may be exploited to provide systematic guidance to the search, so that solution diversity is preserved in a 'natural' way [37]. An approach on this issue is as follows.

The formation of the Pareto front is based on objective functions of the weighted N objectives $f_1, f_2, \ldots, f_N$ which are of the form

$$F_i(\boldsymbol{x}) = f_i(\boldsymbol{x}) + \sum_{j=1, j\neq i}^{j=N} a_{ji} f_j(\boldsymbol{x}), i = 1, 2, \ldots, N \tag{4.3}$$

where $F_i(x)$ is the new objective function; a_{ji} is the designated amount of gain in the j-th objective function for a loss of one unit in the i-th objective function. Therefore the sign of a_{ji} is always negative. The above set of equations requires fixing the matrix a. This matrix has all ones in its diagonal elements. To find the Pareto front of a maximization problem we assume that a solution parameter vector $\boldsymbol{x}_1$ dominates another solution $\boldsymbol{x}_2$ if $F(\boldsymbol{x}_1) \geq F(\boldsymbol{x}_2)$ for all objectives. At the same time a contingent equality is not valid for at least one objective. $F_i(x)$ functions define the contour lines, which form a convex hull with the coordinate axes.

In the case the contour lines are horizontal and vertical, then a_{ji} becomes zero and $F(\boldsymbol{x}) = f(\boldsymbol{x})$.

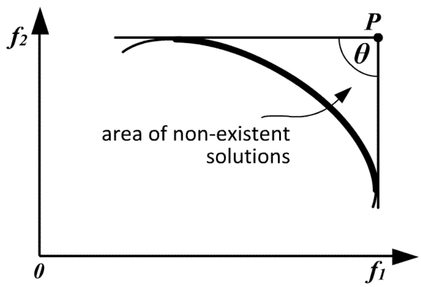


Fig. 4.9 Contour lines defining the dominated region with respect to *P:* greedy dominance

Explicitly we write

$$F_1(\boldsymbol{x}) = f_1(\boldsymbol{x})$$
$$F_2(\boldsymbol{x}) = f_2(\boldsymbol{x}) \quad (4.4)$$

Figure 4.9 shows the contour lines corresponding to two linear functions for a two-objective MO case. In this case the objectives are in particular subject to maximization. From the figure it is important to note that the point P is ultimately subject to identification as an ideal solution. That is, it is the explicit ultimately attainable goal defined by the boundaries of the objective functions. This point is referred to as *ideal point*. The point P can be sought by means of appropriate methodologies, one of which is the method of genetic algorithm applied in this approach. For the search process itself, different strategies can be followed. Below, two strategies are described for the sake of providing insight into the search process of the multi objective optimization approach applied in this work.

In one strategy the point P denotes the ideal point. The premise of the strategy is that beyond this point the solution is not acceptable or meaningless due to the limits of the objective functions. The algorithm using the orthogonal lines is called *greedy MO algorithm*. The modification of the contour lines departing from horizontal and vertical ones defines a modified solution space. This is shown in figure 4.10 by hatched areas. These areas are termed *as domain of relaxation* in this thesis. Solutions found in these areas are not valid Pareto solutions, although they seemingly are solutions. For this modified solution space the area of non-existent solutions in the convex hull is diminished. This is at the expense of deviating from the strict non-dominated solution condition as shown in figure 4.9. From figure 4.10 it should be noted that in case the strict non-domination condition is relaxed, the Pareto front is smaller compared to the case of greedy search in figure 4.10. However, this might be compensated by moving the Pareto surface forward more with pressure, so that the front comes closer to the point P. This strategy in the parameter space allows selection of the parameters in such a way that the relaxation domains in the objective functions space come to existence.

The relaxed dominance approach in figure 4.10 appears to be similar to the *reference-point based approach* [34]. However, comparing both approaches it should be noted that in the relaxed dominance approach proposed put forward in this thesis, a reference point is not used. This is in contrast to the reference-point based approach. Figure 4.10 and its ideal point *P* serve as a conceptual explanation of the Pareto-optimality. In particular the difference between the relaxed dominance compared to the greedy dominance concept is addressed. Namely, concerning the point *P*, by making the angle θ larger than 90 degrees the area of non-existent solutions is reduced compared to the greedy case. Therefore the Pareto front is allowed to establish closer to the ideal point *P*, while at the same time the front is expected to be more diverse. In figure 4.10 it is seen how the greedy front comes closer to the Point *P* through the widening of the angle θ. This is indicated by means of arrows. In the relaxed approach, the avoidance of aggregation to some extend is due to the modification of the objective space, where the space becomes larger, and thus the density of solutions per unit length along the front is expected to become lower.

In a second strategy a hypothetical point designated as *P'* denotes the explicit predefined sub-attainable goal. This goal is positioned somewhere in the convex hull, and is hopefully not far from the point *P*. This is shown in figure 4.11. It is to note that, since the point *P* is not explicitly known, the position assessment of the point *P'* is a problematic

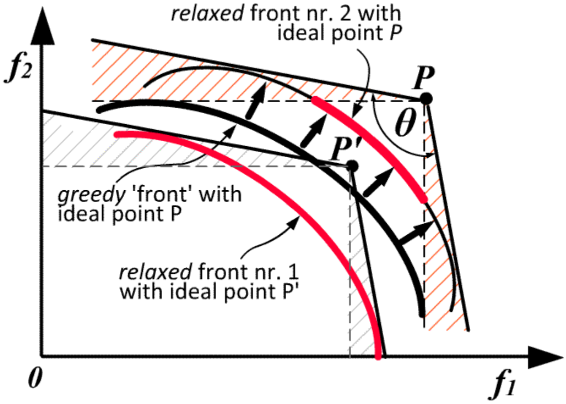


Fig. 4.10 Contour lines defining the dominated region with respect to *P:* relaxed dominance

issue. In any case P' is a sub-optimal point implying some form of relaxation in the search process. In this case, the premise of the strategy is that the increased size of the area of non-existent solutions for orthogonal search domains is compensated with the increase of the size of the Pareto front in the relaxed case. At the same time the area of non-existent solutions is reduced for relaxed search domains. This is seen from the figure 4.11, where Pareto fronts having orthogonal and non-orthogonal contour lines are shown together. The latter strategy in the parameter space allows selecting the parameters in such a way that the relaxation domains in the objective functions space of figure 4.10 do *not* come to existence. However, P' may come reasonably close to P, while this may not happen too. In the latter case still a Pareto front can be obtained. However the front remains to be poor.

Due to this very reason, it is noteworthy to point out that in effective multi-objective optimization to obtain a Pareto surface is only half of the task to be accomplished. The other half is to obtain an effective front at the same time. The second strategy may not allow the parameters to explore the whole solution space due to the non-convex region defined by orthogonal and non-orthogonal axes in figure 4.11. From the figure we note that as the degree of relaxation increases, the diversity of solutions along the Pareto surface increases, too.

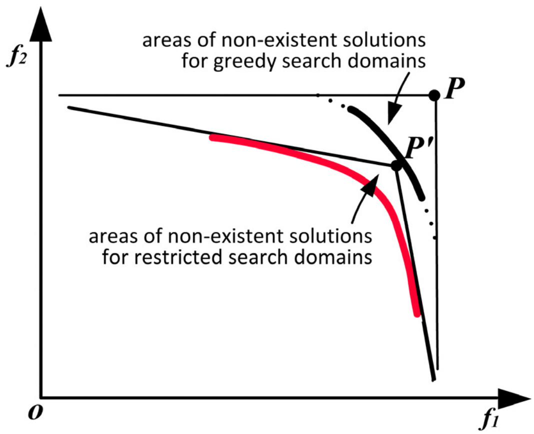


Fig. 4.11 Contour lines defining the dominated regions in orthogonal versus relaxed search

Although the Pareto front concept is well defined, the formation of this front is dependent on the implementation of the MO algorithm. Also it depends on the nature of the application. One interesting application is *the cone domination* approach [37], where the a_{ij} coefficients in Eq. (4.3) are determined in advance, so that the optimality search is carried out in the *F*-space in place of *f*-space. However, the necessity to modify the coefficients during the search makes the approach rather impractical in the higher dimensions.

Based on the strategy depicted in figure 4.11 the search process can be relaxed in a controlled way to increase solution diversity and at the same time to have effective multi-objective optimisation. That is, the mechanism pushing the Pareto front forward is well balanced. This is a novel approach and it is explained below.

For the *greedy* application of the MO algorithm, one uses the orthogonal contour lines at the point P as shown in figure 4.12. In this figure the point P denotes one of the individuals among the population in the context of genetic algorithm based evolutionary search. In the Greedy search many potential favourable solutions are prematurely excluded from the search process.

This is because during the search each solution of the population is represented by the point *P*, and the dominance is measured in relation to the number of solutions falling into the *search domain* within the angle $\theta=\pi/2$. However, please note that during the search this domain is just the opposite and symmetrical to the domain shown also in figure 4.10; the point of symmetry is the solution P. This is seen from figure 4.12. The resulting Pareto front corresponds to a non-orthogonal *search domain* as shown in figure 4.11.

4.4.2 A novel concept for Pareto optimality – relaxed Pareto ranking

To avoid premature elimination of potential solutions, a relaxed dominance concept is implemented, where the angle θ can be considered as the *angle for tolerance* provided $\theta>\pi/2$. The wider the angle is beyond $\pi/2$ the more tolerant the search process is, and vice versa. For $\theta<\pi/2$, θ becomes the *angle for greediness*. Domains of relaxations are also indicated in figure 4.12. From figure 4.10 the differences between the greedy dominance and relaxed dominance methods are seen. In the former case the solutions aggregate. In the latter case, the solutions are more diversified and more effective. In both cases, the fitness of the solutions can be ranked by the fitness function

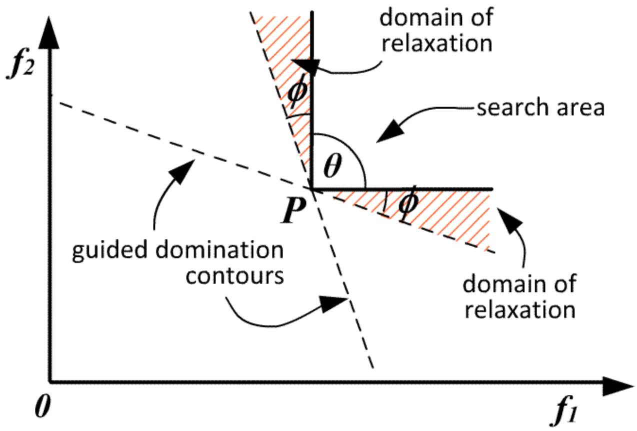


Fig. 4.12 Contour lines defining the search areas

$$R_{fit} = \frac{1}{N(\theta) + n} \tag{4.5}$$

where n is the number of potential solutions falling into the *search domain*. Although $N(\theta)$ can be modified during the search, it is expectedly constant once θ is determined. However, without the analysis of the functionality of $N(\theta)$ it is difficult to establish such a function by experiments.

Above considerations and the ad hoc formulation given by Eq. (4.5) can be put into more precise mathematical terms based on the formulation given by Eq. (4.3), as follows. Let Eq. (4.3) be expressed by

$$\begin{aligned}
F_1 &= f_1 \quad + a_{21} f_2 + ...+ a_{n1} f_n \\
F_2 &= a_{12} f_1 + f_2 \quad + ... + a_{n2} f_n \quad . \\
&.......................... \\
F_n &= a_{1n} f_1 + a_{2n} f_2 + ... + a_{nn} f_n
\end{aligned} \tag{4.6}$$

In matrix equation form, Eq. (4.6) becomes

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \cdots \\ F_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & ..... & a_{n1} \\ a_{12} & a_{22} & ..... & a_{1n} \\ .... & & & \\ a_{1n} & a_{2n} & ..... & a_{nn} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ ..... \\ f_n \end{bmatrix} \tag{4.7}$$

For the sake of simplicity in the description below only two objectives are considered while the results are valid for any dimension. The objective functions for this case are given by

$$\begin{aligned}
F_1 &= f_1 + a_{21} f_2 \\
F_2 &= a_{12} f_1 + f_2
\end{aligned} \tag{4.8}$$

In a two-dimensional coordinate system the contour lines in figures 4.9 and 4.10 are orthogonal and non-orthogonal respectively. The search area in the latter case includes also the domains of relaxation. These are added to the search area of the orthogonal system, as illustrated in figure 4.12.

In the non-orthogonal system the search area for the favourable solutions is wider. At the same time some of the solutions are not dominating the solution at the point ***P*** seen in figure 4.12. However, as a trade-off it provides more diversity at the final Pareto front, while the front is not entirely non-dominated. The solutions at the front are more probably non-dominated in the middle part of the front. This is where f_1 and f_2 are close to each other. Conversely, the solutions may be more dominated at the regions close to edges of the front [38]. This is a property of the *cone dominance* approach. This situation occurs since in the cone domination approach a greedy algorithm is applied using a non-orthogonal system taking the point *P* as origin. By doing so, the search algorithm remains the same but it uses the coordinates of the new non-orthogonal system. However, this cone dominance approach does not address the problem of aggregation. This becomes especially problematic in higher multi-dimensional optimization. This means, the Pareto front is potentially wider in the F-space. However, without resolving the aggregation phenomenon the potentially wider Pareto front remains ineffective.

In contrast to the cone dominance approach, in this thesis each member of the population is considered to be represented by point P seen in figure 4.12, and the solutions falling into the relaxation domains are included to the non-dominated solutions. In other words, some dominated solutions are accrued to the non-dominated ones to form the next-generation solutions. This means, the orthogonal system is not replaced by the non-orthogonal system but the greedy non-dominated orthogonal search space is relaxed. The relaxed domains simply contribute to the greedy search domain with some additional, potentially lucrative solutions.

Interestingly, this situation is similar to the classical gradient-based optimization method where each iteration the step-length towards the global maxima or minima determined by the gradient. The step length should be small enough to ensure the stability of the convergence [39, 40]. If the step length is zero, the approach to minima or maxima does not occur. If it is too big, convergence does not occur. For similar reasons, in the evolutionary computation the angle ϕ defining the relaxation domain should be kept small. In this way the stability of the algorithm is maintained and the effectiveness is enhanced. It should be noted that, although the angle ϕ is small it plays role for each population member at each generation. This makes the net effect of the relaxation highly significant. The role of angle ϕ is inversely comparable to that of the step length parameter of gradient-based optimization. If ϕ is zero, greedy search in orthogonal system occurs. Then the final result in the extreme case aggregates to one solution. If it is too big, convergence does not occur.

In Eq. (4.6) the small-enough designation of the parameters a_{ji} is crucial for the performance of the evolutionary computation. It is characterized by the cosines of the angle between the respective coordinate axes. The angle ϕ is expected to be equal to 10^o or less, although it is also application dependent. The normalization of these cosines yields the directive cosines of the coordinate systems.

The coordinate transformation between orthogonal and non-orthogonal systems is determined by these direction cosines [5]. If we denote the direction cosines as q_{ij}, the transformation matrix becomes

$$Q = \begin{bmatrix} q_{11} \; q_{12} \; \cdots\cdots \; q_{1n} \\ q_{21} \; q_{22} \; \cdots\cdots \; q_{2n} \\ \cdots\cdots\cdots\cdots \\ q_{n1} \; q_{n2} \; \cdots\cdots \; q_{nn} \end{bmatrix} \tag{4.9}$$

which transforms the non-orthogonal system to the orthogonal system and vice versa via

$$x = Q\,x' \tag{4.10}$$

$$x' = Q^{-1}\,x \tag{4.11}$$

where x' denotes the non-orthogonal system $\boldsymbol{x}' = [x_1', x_2', \ldots, x_n']^T$. The directive cosine row vectors of Eq. (4.7) are given by

$$d_i = [d_{1i}\; d_{2i} \;..d_{ni}] = \frac{1}{\sqrt{\sum_{j=1}^{n} a_{ji}^2}} [a_{1i} \;\; a_{2i} \;\; .... \;\; a_{ni}] \tag{4.12}$$

Direction cosine row vector corresponds to column vectors in Eq. (4.9), so that for each column in Eq. (4.9)

$$\sqrt{\sum_{i=1}^{n} q_{ji}^2} = 1 \qquad j = 1,2,....,n \tag{4.13}$$

The direction cosines matrix formed by the respective direction cosine row vectors is related to Q in Eq. (4.9) by

$$D = \begin{bmatrix} [d_{11}\ d_{21}\ \ .....\ d_{n1}\ ] \\ [d_{12}\ d_{22}\ \ .....\ d_{n2}] \\ .... \\ [d_{1n}\ d_{2n}\ \ .....\ d_{nn}] \end{bmatrix} = Q^T \tag{4.14}$$

In a two-dimensional case the directive cosine row vectors are given by

$$\begin{aligned} d_1 &= [d_{11}\ d_{21}] = \left[a_{11} / \sqrt{1 + a_{21}^2} \qquad a_{21} / \sqrt{1 + a_{21}^2}\ \right] \\ d_2 &= [d_{12}\ \ d_{22}] = \left[a_{12} / \sqrt{1 + a_{12}^2} \qquad 1 / \sqrt{1\ \ + a_{12}^2}\ \right] \end{aligned} \tag{4.15}$$

The coordinate transformation of point P in figure 4.12 is given by

$$\begin{aligned} \begin{bmatrix} f_1^P \\ f_2^P \end{bmatrix} &= \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} F_1^P \\ F_2^P \end{bmatrix} \\ &= \begin{bmatrix} d_{11} & d_{21} \\ d_{12} & d_{22} \end{bmatrix} \begin{bmatrix} F_1^P \\ F_2^P \end{bmatrix} \end{aligned} \tag{4.16}$$

and

$$\begin{aligned} \begin{bmatrix} F_1^P \\ F_2^P \end{bmatrix} &= \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}^{-1} \begin{bmatrix} f_1^P \\ f_2^P \end{bmatrix} \\ &= \begin{bmatrix} d_{11} & d_{21} \\ d_{12} & d_{22} \end{bmatrix}^{-1} \begin{bmatrix} f_1^P \\ f_2^P \end{bmatrix} \end{aligned} \tag{4.17}$$

The importance of the coordinate transformation becomes dramatic especially in higher dimensions. In such cases the spatial distribution of domains of relaxation becomes complex and thereby difficult to implement. Namely, in multidimensional space the volume of a relaxation domain is difficult to imagine. And more importantly it is difficult to identify the population in such domains. Therefore one needs a systematic approach for identification by computation and not by inspection or anything else. This systematic approach is the coordinate transformation as follows. Basically for each solution point, say P in figure 4.12, the point is temporarily considered to be a reference point as origin. All the other solution points in the orthogonal coordinate system are converted to the non-orthogonal system coordinate by Eq. (4.11).

For instance for three objectives, we write

$$F = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} 1\ a_{21} & a_{31} \\ a_{12}\ 1 & a_{32} \\ a_{13}\ a_{23} & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}, \tag{4.18}$$

where three parameters should be designated, namely a_{12}, a_{13}, and a_{23} in advance if the coefficient matrix is taken to be symmetrical. They are expectedly around the value -0.1

and the coefficients can be modified during the search, if necessary. The presence of the relaxation guarantees the prevention of aggregation as it is meant for.

After conversion, all points which have positive coordinates in the non-orthogonal system correspond to potential solutions contributing to the next generation in the evolutionary computation. If any point possesses a negative component in the new coordinate system, the respective solution is not counted to dominate *P*. This is because otherwise such a solution may lead the search in a direction away from P. This fundamental process employing Eq. (4.11) is very easy to implement in two-dimensional space. However, in higher dimensions it is extremely complex and this is the reason of motivation to develop many different methods for effective Pareto front formation in the literature [41-43].

Here one remaining insignificant question is the selection of the non-orthogonal system. This is in general application dependent. However, the domain of relaxations should be small compared to the total search domain implying small angles between the respective pair of coordinate axes. The angles should be kept to smaller values as the multi-dimensionality increases. Otherwise some unfavourable domains seemingly and erroneously appear to be favourable yielding poor Pareto front formation.

4.5. Fuzzy neural tree (FNT) & multi-objective evolutionary algorithm (MOEA) = Intelligent Design Objects (IDO)

4.5.1 Components forming intelligent design objects

Most design criteria are soft to some extend. This is due to the conflicting nature of the requirements. Examples of such criteria are visual perception aspects and requirements for 'high functionality'. Due to the softness fuzzy neural trees (FNTs) presented in the previous chapter are appropriate models to evaluate the performance of alternative designs.

Using such a model we can equip the multi-objective design generation process with human-like reasoning capabilities. This means the genetic search makes use of our knowledge during its search process. This way that we can evaluate the designs with respect the criteria, and arrive at solutions that satisfy them. This coupling of a neuro-fuzzy performance evaluation model and genetic design generation is shown in the figure 4.13. In the figure the components belonging to design generation are shown in an orange box, and the components belonging to performance evaluation are shown in a green box.

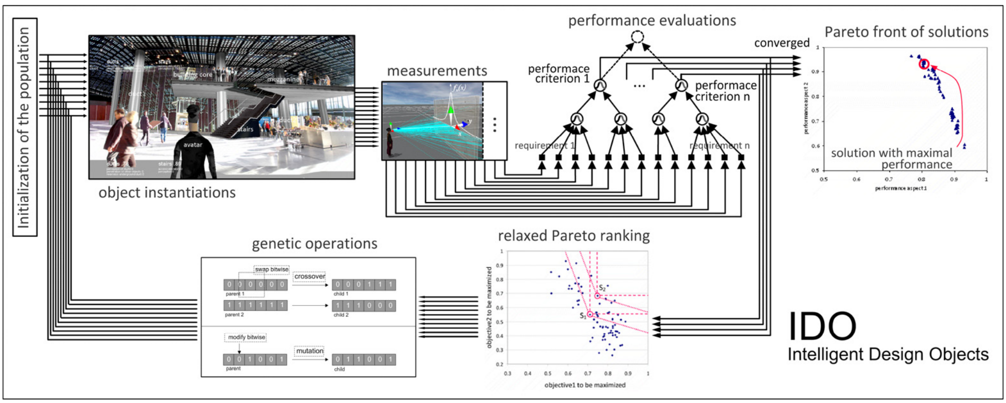


Fig. 4.13 Cognitive system based on intelligent design objects

From figure 4.13 we note that the computational design system starts its processing by generating a population of random solutions within the boundaries we put forward. Then certain properties of these solutions are measured, such as positions, perceptual properties, energy consumption properties and the like. In the present work the focus is on perceptual properties, because they are of general relevance and there are no models existing to measure such properties, yet. The information obtained by means of the measurements is fed into the model for performance evaluation. There the measured values are converted into degrees of satisfaction for their associated design requirements. For example the degree of perception an object in the design has is checked against the desired value. Based on an associated fuzzy requirement function at the terminal node the corresponding degree of satisfaction is obtained. This is done for all elemental requirements and the information regarding their satisfaction is processed through the neural model. This yields the degree of performance at the outputs on the penultimate level of the neural tree model.

The values at these outputs are then used in order to rank the solutions regarding their degree of Pareto dominance. In order to maintain diversity during the genetic search, the solutions are graded using the relaxed dominance concept described above.
Thereafter the genetic operations are performed. This way a new population of solutions is formed. The new solutions are likely to satisfy the criteria more, i.e. the outputs at the penultimate node are higher. This process is repeated until the generation process converged to a diverse Pareto front of solutions.

4.5.2 Cognitive behaviour of the intelligent objects

In the initial phase of design we are usually unable to specify the importance among criteria. This is because we are unable to foresee the implications of such a specification on corresponding solutions. This means the weights on the penultimate level of the tree are not specified a priori. Therefore, in the multi-objective case the Pareto solutions are deemed to be equivalently valid. This means that they may not be distinguished unless second-order criteria are brought into play. Due to the particularities of the neural tree concept being used for performance evaluation the selection of potential second-order preferences is accomplished systematically. In particular the aim is to distinguish the solutions with respect to their *maximal attainable performance* provided the weights on the penultimate level may be freely selected. That is, for any possible preference a designer may have regarding the criteria, we are looking for the maximum performance score a certain solution is able to reach. Then we select the solution with the highest scores

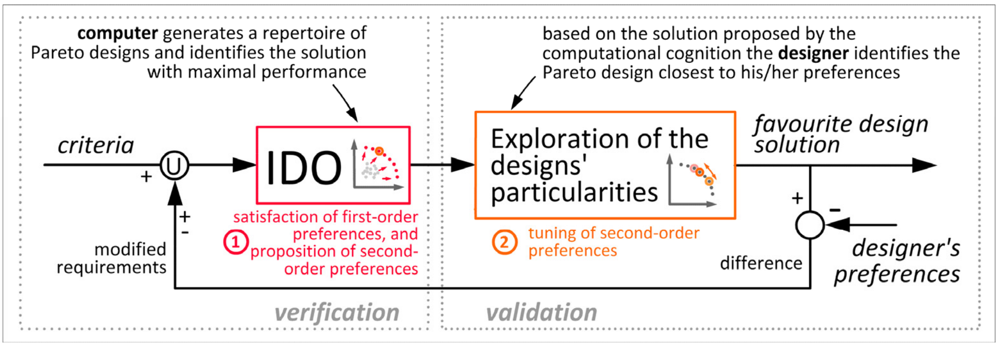


Fig. 4.14 Cognitive design approach

among all of them. This solution has the *maximal performance* as defined in this thesis. This identification is accomplished as follows.

At the root node of the neural tree, the performance score is computed by the defuzzification process given by

$$w_1(1)f_1 + w_2(1)f_2 + ... + w_n(1)f_n = p, \tag{4.19}$$

where $w_1+w_2+...+w_n=1$; f_1-f_n are the outputs at the penultimate nodes; p is the design performance, which is requested to be maximum. The node outputs f_1 - f_n can be considered as the *design feature vector **f***. The reflection of these features at the design performance can be maximally performed if the weights w_1-w_n define the same direction as that of the feature vector. The vector ***w*** containing the weights on the penultimate level is termed priority vector.

Normalising the components and equating them to the weights yields

$$w_1 = \frac{f_1}{f_1 + f_2 + ... + f_n}\ ; \quad w_2 = \frac{f_2}{f_1 + f_2 + ... + f_n}\ ; \quad ...\ ; w_n = \frac{f_n}{f_1 + f_2 + ... + f_n} \tag{4.20}$$

Due to Eq. (4.20) the performance given by (4.19) becomes

$$p = \frac{f_1^2 + f_2^2 + ... + f_n^2}{f_1 + f_2 + ... + f_n} \tag{4.21}$$

Above computation implies that, the performance p for each genetic solution is uniquely determined. Therefore, (4.21) is computed for all the design solutions on the Pareto front. Then the solution with *maximal performance* is selected having a characteristic priority vector $\boldsymbol{w}^*$.

Whenever the intelligent system generates a solution that outperforms any previous solution regarding its *maximal performance*, the design objects assume the properties of the new best design. That is, the objects move in a smooth transition from their previous locations to the new locations, and they change their colour gradually from their previous colour to the new colour. The smoothness of the transition is done, so that a designer can easily follow the behaviour of the objects, i.e. the progress of the generation process.

This systematic change of properties of objects with certain complex objectives leads to the interpretation that the object movements are activities deemed to be intelligent. Therefore, such design objects are termed intelligent design objects (IDO). This means the objects exhibit human-like behaviour to some extent. The human-like behaviour is that the objects have an associated reasoning process to maximize their performance in accordance with the requirements. Naturally, one can tend to think that the object movements are due to some feedback mechanism, so that the net-activity is seemingly adaptive rather than intelligent. However, the situation is not that simple since the search for suitable designs yields voluminous information subject to processing, while the goal of this processing is not unilateral. That is, there are several objectives to achieve while these objectives are in conflict in general.

Beyond the intelligent behavior it is noted that this selection among the solutions using Eq. (4.21) is a form of *machine cognition*. This is because the selection of a solution involves determining the relative importance among the criteria as ***w****. This is an act where second-order preferences are in the play. Namely these preferences were not included in the criteria, but they become apparent through the process of generating Pareto optimal solutions. It is noted that the ***w**** vector is changing throughout the search process, whenever the search finds a new solution with maximal performance. This can be interpreted as a process where the intelligent objects develop their cognition throughout the search. The development is regarding which preferences among the criteria are most likely to yield maximal design performance in the face of the present design requirements and solution space.

Concerning the definition of an intelligent design object, the term object refers to a set of parameters specifying the specific condition of the object both in terms of its physical properties as well as its abstract qualities. For example a design object like a table has with and height parameters specifying its physical properties, and at the same time it has associated attributes such as functionality, aesthetics, stability etc. Design object as understood in this work thus refers the entire set of attributes. A whole scene consisting of a number of building elements may be considered as a single intelligent design object. Depending on the task it may be convenient to subdivide the scene into sub-objects, where every sub-object has its unique associated performance criteria. That is, objects may operate independently from other objects in the sense that they influence different performance criteria. However, in case some parameters of one entity in a design influence the performance of another one then both entities together form an intelligent design object. This is because in the intelligent processing all parameters that influence an object's performance are considered together, so that the most suitable *combination* of parameter values is identified.

It is emphasized that the boundaries for the evolutionary search should be selected in accordance with the type of object considered, as well as the particular design context. For example, say we wish to place stairs to connect from a floor to a mezzanine, while the mezzanine's position is also subject to evolutionary search. In this case it is clear that the search for a suitable position of the stairs does not include locations where the stairs are not connected to the mezzanine at all. In the intelligent object approach this basic specification is invariantly taken into account in the search by means of dynamically adapting the search boundary belonging to an object. This feature is a significant issue making the evolutionary search an intelligent act as opposed to being merely adaptive.

It is noted that intelligent design objects may be employed in a context of conventional design objects, i.e. objects without associated intelligent or cognitive behaviour. For example, when an object's features are already determined and not subject to modification the object is not equipped with intelligent behaviour. This is the case for instance when the object is part of an existing spatial situation. However, an IDO considers its relation with conventional objects, when this relation is significant with respect to the IDO's performance.

4.5.3 A cognitive approach for performance-based design

Using Eq. (4.21) a solution with maximal performance is identified among the Pareto optimal solutions. This solution has a specific priority vector **w***, i.e. specific weights on the *penultimate level* in the neural tree.

When we inspect this solution we may find that it does not entirely coincide with our preferences. This is expected, since it would be a rare contingency if the machine cognition would entirely match our own cognition. In this case the solution with maximal performance does not reflect certain features we particularly wish to have strongly enough in the design. For example we may find the design does not provide enough visual privacy, or it is not functional enough, or anything else.

With such an assessment we are able to increase the component of the priority vector **w** that belongs to the feature we wish to boost. Based on the new priority vector the intelligent objects present another design among the Pareto solutions that matches the new preference vector most closely. It is important to note that the new solution will be the solution where our increase in privacy yields the *least possible sacrifice of performance* on the other criteria. This is inherently ensured due to the Pareto front concept. The exact implications of the new priority vector we give are difficult to foresee. That is, it is uncertain how much to increase the privacy to arrive at a solution that suits better to our preferences. Therefore the exploration on the front is generally an iterative process. This is shown in figure 4.14. Inspecting the new solution we are in a position to judge whether it matches our preferences, and modify w accordingly. This process is repeated until the most suitable solution among the Pareto solutions is found. We term this process 'exploring the Pareto front.' It is noted that during the designer's exploration on the front the behaviour of the

objects is intelligent and not cognitive. That is the designer's cognition takes over in this phase. Whereas in the initial act of identifying the solution with maximal performance and its corresponding priority vector ***w****, machine cognition was active.

It is noted that this alternating sequence of involvement of human and machine cognition is a unique form of communication with a machine. Based on the machine's cognitive proposal we are able to 'ask' the intelligent objects to show us a solution that has 'more privacy' for instance, while we sacrifice as little as possible regarding the other criteria. This form of human-machine interaction is normally unthinkable since second-order demands, such as 'more privacy' can only be communicated among humans, due to the vagueness of the concept. The interaction between designer and an intelligent system involving machine cognition is termed as *cognitive design approach*. It is shown schematically in figure 4.14.

The benefit using the computational cognition as starting point for our exploration on the front is that it allows us to search the potentially vast amount of Pareto solutions efficiently. This is because it provides the solution with maximal performance in an unbiased sense, i.e. although we did not specify relative importance among objectives, yet. The solution with maximal performance using Eq. (4.21) is the appropriate starting point when there is no information on the priority vector. Also it indicates the maximal performance that may be achieved based on the present criteria. This means it gives information how much performance score we are sacrificing when we select a design with a priority vector different from ***w****. This way the performance of other solutions is put in perspective. The latter feature is relevant in particular when we are still not entirely satisfied with the most suitable Pareto solution we found after exploring the Pareto front. The latter solution is termed as favourite solution. In this case it is clear that the Pareto front is not suitable for our ultimate goal, so that we should modify the criteria. This is seen from figure 4.14 by means of the feedback loop leading from the output of a favourite solution back to the criteria. Having seen the Pareto solutions we are able to modify the requirements based on the novel insight we gained through the exploration on the Pareto front. That is, we gained understanding on where deficiencies of the design are, since we were able to experience the implications of our current criteria. In particular we were able to gain an overview regarding the conflicts among our criteria, and we realize the role of certain solution parameters play more clearly. This way we are able to adjust design criteria consciously, so that they match closer to our preferences. That is, we are more conscious to adjust criteria to introduce new requirements, adjust the tolerances for satisfaction at the terminal level of the tree, or modify the relative importance among aspects, and then let the computational process run again based on the improved criteria.

This process of modifying the criteria is termed as validation in the literature and it is recognized that it is a problematic issue to systematically support validation [44]. The cognitive approach provides a novel computational means to support this process. The process of verification/validation is iterated until the favourite solution matches our preferences entirely.

It is emphasised that the cognitive aspect of the approach is essentially due to two factors. One factor is the computational cognition proposing second-order preferences that appear most promising for the present task. A second factor is the designer's exploration of the particularities among the solutions on the Pareto front comparing his own cognitive judgment of the task with the result from machine cognition. It is expected that this will be necessary several times during a design process. It is also emphasized that this process is an interactive process where machine and human cognition are interfaced via the intelligent objects. This interactive process may therefore be termed *Interactive Intelligent Design Objects* (I^2DO). It is to be noted that this approach does not claim provision of a fully automated design system. It always involves a decision maker in the loop for interactivity. In this respect the instantiation of the design objects while the decision maker explores the Pareto front is a significant feature facilitating the interaction. Namely the designer is able to judge the suitability in parameter space, i.e. facing the concrete implications of his/her decision, while manipulating the relative significance of abstract dimensions, i.e. the objective space.

4.6. Discussion

4.6.1 A formalism for design

Based on the above considerations the cognitive approach described may serve as a formalism for architectural design.

An architectural design is expected to provide a number of desirable features. These range from satisfying basic and higher-level human needs, to creation of cultural value. When we are designing, due to the multiplicity and soft nature of the design criteria we are usually not fully aware of our preferences regarding the desirable features the design should possess [6]. Exploring solution options in a goal-oriented manner during design, we bring our awareness to a higher level. This means we get to know what the implications are of the criteria we are taking into account. In the computational approach presented in this thesis this higher awareness is modelled by means of Pareto front of solutions. It is established based on simultaneously maximizing the most abstract performance aspects characterizing the design, such as functionality, perception aspects, etc. Based on this awareness we are able to access with greater precision to preferences we were initially vaguely aware of. Most appropriate second-order preferences become apparent to us only at the time we face solution alternatives. This is because the conflicts among our criteria are difficult to grasp beforehand, so that their impact on the solutions are difficult to foresee. In the present approach second-order preferences are modelled by means of a weight vector associated with the penultimate node connections of the neural tree. This is to select solutions on the Pareto front that fit both our first-order criteria, and also our second-order preferences. The involvement of a second-order selection process makes design a cognitive act, beyond being merely an effort to maximize a single, albeit complex, performance criterion.

It is emphasized that the cognitive process occurring naturally in the mind of a designer is supported in the present approach by *machine cognition*. This is accomplished by letting the computer establish a Pareto front of solutions, and selecting the maximally performing solution among them. It is emphasized that the computer performs this selection based on the contingent availability of solutions, i.e. its preference for a certain solution is not due to a-priori fixed preference for design features, but this preference emerges during the search for optimality. Concisely the design formalism based on the cognitive approach is:

Design is a search for Pareto optimality with the ultimate goal to satisfy soft first-order criteria as well as emerging second-order preferences. From this it is clear that design is a cognitive process.

This formalism is novel in several respects, as this is explained below.

4.6.2 The cognitive approach in perspective

Although the potential of evolutionary computation for architectural design has been recognized [45, 46], it was not clear up till now how such a method should be expanded so that it is able to model the complexity of the design. These limitations refer in particular to the computational treatment of the soft issues of architectural design, i.e. the linguistic nature and the multiple attribute relations inherent to the design criteria. In this respect the particularities of the special neuro-fuzzy computation involved in the novel intelligent object approach play a significant role. Due to these particularities the Pareto concept is revealed as essential concept behind a theory of design.

Another factor that hindered the development of computational cognitive approaches for design is that cognition is not very well understood up till now [47-50]. This is because the cognition process itself is difficult to model based on observations of a designer's activity. Human cognition is much more complex than the observable manifestations that occur when we are designing. This is clear when we consider that cognition involves highly complex brain activities. The starting point of a design is always different [51]. Due to the complexity of a design, the path a designer is taking is sensitive to this starting condition and unique. Therefore, conclusions about design that are based upon such observations can hardly be validated, and consensus is missing as to what constitutes design exactly. This is indicated by the multitude of descriptions on what design is. These range from *decision making* [52], *problem solving* [53], *application of pattern languages* [54] and *execution of shape-rules* [55] to *intentional action* [56], *reflection-in-action* [57], *situated social interaction* [58], *thinking* [59], *anticipation* [60]. These descriptions are all valid, however due to the soft nature of the design process it is generally difficult to derive computational methods and formalisms based on them.

Cognition is understood in this thesis as the faculty to take a decision based on an awareness of solution alternatives that are all fulfilling certain primary objectives. The feature making the process cognitive is to select among these apparently equivalently valid solutions based on second-order preferences that were not included in the establishment of the solutions. Such second-order preferences are due to the particularities among the alternative solutions which cannot be foreseen prior to designing. Cognition is closely linked with intelligence as intelligence is the machinery able to establish the awareness of these solution alternatives [61]. Cognition identifies the most suitable solution based on this awareness.

With this understanding cognition is a means to handle the contingency of architectural design tasks effectively. In other words the complexity of decisions is partially addressed through intelligent processing dealing with the general relations among environmental information. After that contingency-based issues are taken care of by exercising second-order preferences and selecting among the potential options. In this way the effectiveness to deal with complex tasks where there is no generally valid solution recipe is enhanced.

However, it is noted that *not every* multi-objective approach with ensuing preferential selection process is to be considered as a cognitive approach. The cognitive component is especially due to the human-like reasoning executed in the fuzzy neural model for performance evaluation. The model involves abstract performance concepts, such as functionality, perception aspects, etc. In this respect the transparency feature of the fuzzy neural tree is significant, as it allows treating soft, i.e. linguistic design criteria in the context of a multi-objective optimization problem. Especially computational cognition in the framework of multi-objective optimization requires the use of the fuzzy logic approach put forward in this thesis. This is because seemingly incompatible concepts are brought into the same domain for integral considerations. Namely, based on this the computer is able to compare solutions regarding their maximal performance, although no prior commitment was made regarding the priorities among criteria.

The resulting benefit for an architect applying the approach is that he/she can be more certain he/she did not miss a superior solution. An architect has unique experience and intuition. Based on the Pareto front of solutions he/she is able to exercise these capabilities with great awareness responding to circumstantial influences. This way the complexity of design tasks can be faced with computational means, so that a designer is able to make use of his capabilities more effectively than conventionally it is the case.

It may be noteworthy to mention that computational design is by no means a 'competition' for a designer; it is a means to elevate the quality of the result by providing the designer with sophisticated means to exercise his knowledge, and intuition more effectively.

4.7. Conclusions

A novel approach for performance-based design generation is presented. The method generates designs that fulfil the design criteria an architect put forward. It takes into account the vague, linguistic character of design requirements, and the large amount of possible solutions. This means it does not require that we specify the relative importance among the criteria in advance. The method is based on integrating the performance evaluation described in the previous chapter into a computational process that systematically generates designs with suitable performance. This means the process generates a large number of alternative designs, and then evaluates their performance based on our criteria to generate more favourable solutions in the next processing cycle. In this way, after some time the method generates designs with maximal performance.

Since the importance among the performance aspects is usually unknown, the generation process establishes a collection of solutions, and not merely a single best solution. This means the process is a multi-objective process as opposed to a single-objective process. This is done in order to obtain a variety of equivalently valid alternatives that are all fitting to our criteria, so that we are able to select among them with great awareness. To allow efficient selection the computer proposes a certain solution among the Pareto optimal ones, which has maximal performance. The maximal performance is in the sense that the solution outperforms the other solutions when all possible preferences for the given criteria are exercised without bias.

The essential benefit of the method is that the complexity of design tasks can be handled with powerful computational means, so that a designer is able to concentrate his mental effort on the essential questions with great awareness. In particular in the present approach the three components described in the previous chapters are combined, so that soft requirements, including perception-based requirements, are effectively integrated into the design process. We are able to bring our higher-order preferences into play, when we have Pareto optimal solutions that correspond to the criteria we put forward. Due to the Pareto front the designer is no longer loaded with the task to bear in mind the complex relation between parameters and the performance aspects of the design. This is because this relation is 'embodied' in the form of the Pareto optimal solutions already. This means we are able to pay attention to the issues that were not included in the process identifying the potential solutions, being otherwise hardly aware of these issues.

This chapter concludes the theoretical part of the thesis. It is followed by three applications of the novel approach.

4.8. References

1. Ciftcioglu, Ö., Bittermann, M.S.: Solution diversity in multi-objective optimization: A study in Virtual Reality. World Congress on Computational Intelligence WCCI 2008, Hong Kong (2008)
2. Ciftcioglu, O., Bittermann, M.S., Sariyildiz, I.S.: Building performance analysis supported by GA. In: Tan, K.C., Xu, J.X. (eds.): 2007 IEEE Congress on Evolutionary Computation. IEEE, Singapore (2007) 489-495
3. Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, Ö.: Performance-based Pareto optimal design. In: Rusák, I.H.a.Z. (ed.): TMCE 2008, Izmir, Turkey (2008)
4. Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, Ö.: Multi-objective optimization in the construction industry. AEC 2008, Antalya, Turkey (2008)
5. Ciftcioglu, O., Bittermann, M.S.: From perceptual towards cognitive robotics in the framework of evolutionary computation. In: Pennacchio, S. (ed.): Emerging Technologies, Robotics and Control Systems. International SAR, Palermo, Italy (2009)
6. Lawson, B.: How Designers Think. Butterworth-Heinemann, Oxford (1997)
7. Hamel, R.: Over het Denken van de Architect. AHA books, Amsterdam (1990)
8. Love, P.E.D., Gunasekaran, A.: Concurrent engineering in the construction industry. Concurrent Engineering: Research and Applications **5** (1997) 155-162
9. Deb, K.: Multiobjective Optimization using Evolutionary Algorithms. John Wiley & Sons (2001)
10. Goldberg, D.E.: Genetic Algorithms. Addison Wesley, Reading, Massachusetts (1989)
11. Bellmann: Adaptive control processes: A guided tour. Princeton University Press, Princeton (1961)
12. Huang, W., Lam, H.: Using genetic algorithms to optimize controller parameters for HVAC systems. Energy and Buildings **26** (1997) 277-282
13. Castilho, V.C.d., Debs, M.K.E., Botelho, C.: Using a modified genetic algorithm to minimize the production costs for slabs of precast prestressed concrete joists. Engineering Applications of Artificial Intelligence **20** (2007) 519-530
14. Camp, C., Pezeshk, S., Cao, G.: Optimized design of two-dimensional structures using a genetic algorithm. Journal of Structural Engineering **124** (1998) 551-559

15. Wang, B., Chen, J.: Application of genetic algorithm for the support location optimization of beams. Computers and Structures **58** (1996) 797-800
16. Jo, J., Gero, J.: Space layout planning using an evolutionary approach. Artificial Intelligence in Engineering **12** (1998) 163-176
17. Damsky, J., Gero, J.: An evolutionary approach to generating constraint-based space layout topologies. In: Junge, R. (ed.): CAAD Futures 1997. Kluwer Academic Publishing (1997) 855-874
18. Monks, M., Oh, B.M., Dorsey, J.: Audioptimization: goal-based acoustic design. IEEE Computer Graphics and Applications **20** (2000) 76-90
19. Gero, J., Kazakov, V.: Evoloving design genes in space layout problems. Artificial Intelligence in Engineering **12** (1998) 163-176
20. Nassif, N., Kajil, S., Sabourin, R.: Evolutionary algorithms for multi-objective optimization in HVAC system control strategy. NAFIPS '04 (2004) 51-56
21. Shea, K., Sedgwick, A., Antonuntto, G.: Multicriteria optimization of paneled building envelopes using ant colony optimization. Intelligent Computing in Engineering and Architecture. Springer Verlag, Berlin Heidelberg (2006) 627-636
22. Caldas, L.G., Norford, L.K.: Genetic algorithms for optimization of building envelopes and the design and control of HVAC systems. ASME Journal of Solar Energy Engineering **125** (2003) 343-351
23. Caldas, L.: GENE_ARCH: An evolution-based generative design system for sustainable architecture. In: Smith, I.F.C. (ed.): Lecture Notes in Artificial Intelligence Springer-Verlag Berlin Heidelberg (2006) 109-118
24. Wang, W., Rivard, H., Zmeureanu, R.: Floor shape optimization for green building design. Advanced Engineering Informatics **20** (2006) 363-378
25. Caldas, L.: Generation of energy-efficient architecture solutions applying GENE_ARCH: An evolution-based generative design system. Advanced Engineering Informatics **22** (2008) 59-70
26. Coello, C.A.C., Veldhuizen, D.A., Lamont, G.B.: Evolutionary Algorithms for Solving Multiobjective Problems. Kluwer Academic Publishers, Boston (2003)
27. Fonseca, C.M.: An overview of evolutionary algorithms in multiobjective optimization. Evolutionary Computation **3** (1995) 1-16
28. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Grunert, V.: Performance assessment of multiobjective optimizers:an analysis and review. IEEE Trans. on Evolutionary Computation **7** (2003) 117-132
29. Coello, C.A.C.: An updated survey of GA-based multi-objective optimization techniques. ACM Computing Surveys **32** (1999) 109-143
30. Sato, H., Aguirre, H.E., Tanaka, K.: Controlling dominance area of solutions and its impact on the performance of MOEAs. Springer, Heidelberg (2007)
31. Goldberg, D.E., Richardson, J.: Genetic algorithms with sharing for multimodal function optimization. In: Grefenstette, J.J. (ed.): Second Int. Conf. on Genetic Algorithms. Lawrence Erlbaum Assoc. (1987) 41-49
32. Fonseca, C.M., Fleming, P.J.: Genetic algorithms for mulit-objective optimization: Formulation, discussion and generalization. Fifth Int. Conf. on Genetic Algorithms. Morgan Kauffman, San Mateo (1993) 416-423
33. Fonseca, C.M., Fleming, P.J.: Multiobjective optimization and multiple constraint handling with evolutionary algorithms-Part II: Application example. IEEE Trans. Syst., Man, Cybern **28** (1998) 38-47
34. Deb, K., Sundar, J., Bhaskara, U., Chaudhuri, S.: Reference point based multi-objective optimization using evolutionary algorithm. Int. J. Comp. Intelligence Research **2** (2006) 273-286
35. Ciftcioglu, Ö., Bittermann, M.S.: Solution diversity in multi-objective optimization: A study in virtual reality. Wold Congress Comp. Intelligence, Hong Kong (2008)
36. Tan, K.C., Xu, J.-X.: 2007 IEEE Congress on Evolutionary Computation, Singapore (2007)
37. Branke, J., Kaussler, T., Schmeck, H.: Guiding multi-objective evolutionary algorithms towards interesting regions. Technical Report TR No. 399, . AIFB University of Karlsruhe, Germany (2000)

38. Deb, K., Zope, P., Jain, A.: Distributed computing of pareto-optimal solutions with evolutionary algorithms. In: Fonseca, C., Fleming, P., Zitzler, E., Deb, K. (eds.): 9th annual conference on Genetic and evolutionary computation ACM New York, London, England (2003) 532-549
39. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming. John Wiley & Sons, Inc, New York (1993)
40. Kuester, J.: Optimization Techniques with FORTRAN. McGraw-Hill College (1973)
41. Ishibuchi, H., Murata, T.: Multi-objective genetic local search algorithm. IEEE International Conference on Evolutionary Computation, Nagoya, Japan (1996)
42. Jaszkiewicz, A.: On the computational efficiency of multiple objective metaheuristics: The knapsack problem case study. European Journal of Operational Research **158** (2004) 418-433
43. Hughes, E.J.: Evolutionary many-objective optimisation: many once or one many? : IEEE Congress on Evolutionary Computation CEC'2005. IEEE Service Center, Edinburgh, Scotland (2005) 222-227
44. Wise, J.A., Hopkin, V.D., Stager, P., (eds.): Verification and validation of complex systems: Human factors issues. NATO Advanced Study Institute - ASI. Springer-Verlag, Berlin, Vimeiro, Portugal (1992) 705
45. Frazer, J.: An Evolutionary Architecture. AA Publications, London (1995)
46. Gero, J.: Computational models of innovative and creative design processes. Technological Forecasting and Social Change **64** (2000) 138-196
47. Ashcraft, M.H.: Cognition. Prentice Hall (2005)
48. Ahle, E., Söffker, D.: A cognitive-oriented architecture to realize autonomous behaviour – part I: Theoretical background. 2006 IEEE Conf. on Systems, Man, and Cybernetics, Taipei, Taiwan (2006) 2215-2220
49. Neisser, U.: Multiple systems: A new approach to cognitive theory. European Journal of Cognitive Psychology **6** (1994) 225-241
50. Turvey, M.T., Shaw, R.E.: Ecological foundations of cognition. Journal of Consciousness Studies **6** (1999) 95-110
51. Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, O.: Perception & Architecture. In: Bekkering, H., Hauptmann, D., Heijer, A.d., Klatte, J., Knaack, U., Manen, S.v. (eds.): The Architecture Annual 2005-2006 Delft University of Technology. 010 Publishers, Rotterdam (2007) 104-109
52. Archer: The structure of the design process. In: Broadbent, G., Ward, A. (eds.): Design Methods in Architecture. MIT Press, Cambridge (1969) 76-102
53. Simon, H.A.: Decision making and problem solving. Research Briefings 1986. National Academy of Sciences, Washington, DC (1986)
54. Alexander, C.: The Timeless Way of Building. Oxford University Press, New York (1979)
55. Mitchell, W.J.: The Logic of Archtecture. MIT Press, Cambridge (1989)
56. Galle, P.: Design as intentional action: a conceptual analysis. Design Studies **20** (1999) 57-81
57. Schön, D.A.: The Reflective Practitioner. Basic Books, New York (1983)
58. Gero, J.: Computational models of creative designing based on situated cognition. 4th conference on Creativity & Cognition ACM, Loughborough, UK (2002) 3-10
59. Ostberg, G.: What goes on when a designer thinks? : Cognition, Communication and Interaction. Springer, London (2008) 486-500
60. Zamenopoulos, T., Alexiou, K.: Towards an anticipatory view of design. Design Studies **28** (2007) 411-436
61. Strube, G., Habel, C., Konieczny, L., Hemforth, B.: Kognition. In: Görz, G., Rollinger, C.-R., Schneeberger, J. (eds.): Handbuch der Kunstlichen Intelligenz. Oldenbourg Wissenschaftsverlag, Munchen (2003) 19-72

5. Applications of Intelligent Design Objects

In the previous chapters a cognitive approach for performance-based design is described. This chapter presents two design applications using this approach. The chapter is based on the publications [1-6].

The applications demonstrate the suitability of the approach for use in architectural design. The applications are treated separate sections. They differ with respect to the scale and type of the architectural context, indicating the generic character of the approach. Both examples involve soft requirements, in particular requirements for visual perception qualities. This exemplifies the capability of the approach to handle soft requirements, which are conventionally considered incomputable. The applications involve multiple requirements, which are conflicting, as this is usually the case in design. The ability to deal with this issue is demonstrated as well.

Both applications are presented in the following sequence of steps: First the design task is outlined. Then the formation of the performance evaluation model is described based on the design criteria pertinent to the task. Then the execution of the intelligent object implementation is described. This is followed by the analysis of the resulting designs.

5.1. Application nr. 1

5.1.1. Design task

In the following design application the intelligent object approach presented in the previous chapter is used. We consider the design of an ensemble of residential housing units. The problem is taken from an actual design case. The task is to find suitable locations of a number of housing units on their respective lots. The site belongs to one of the largest areas in the Netherlands subject to development, named Leidsche Rijn. It is a part of the city of Utrecht. In 2007 the total area had 14.000 inhabitants and the number is rising rapidly. The projected amount of inhabitants is 80.000 people by 2015.

The site considered in the application has a size of about 3600m^2. The streets and lots are provided in advance in this case. The site is shown in figure 5.1. In the figure 20 houses are seen. Three of them are existing, namely *E1, E2,* and *E3*, so that 17 houses are subject to optimal positioning. It is emphasized that in this task we are not concerned with the design of the individual houses, or the organisation of the lot. We consider where to place the buildings on their lot, so that the ensemble has some desirable properties. The houses G_a1-G_a6 and G_b1-G_b4 form two groups of houses, which are situated along a line parallel to the perimeter of the neighbourhood. It is an initial basic choice of the architects to align these houses with respect to each other. This is respected as an architectural premise throughout the implementation. Therefore any computational solution identified will have this property. These houses have a square shaped floor plan of 8m by 8m and they are located along a line at equal distance from each other. The south direction in the situation is towards the street indicated as *Noordelijke stadsas* in figure 5.1. The design shown in the figure is a solution proposed by an urban design office.

The houses that are not subject to positioning are *E1*, *E2* and *E3,* since they are existing buildings. It is noted that the existing buildings do play a role in the assessment of the performance of the designs. All buildings are two storeys high. Houses *E1*, *E2* and *E3* have varying floor plan dimensions and orientations; houses *H1-H7* are 12m long, 8m wide and their longer axis is oriented in east-west direction.

The criteria put forward in the present example are visual perception aspects and requirements on the size of the gardens of every house. In the example the perceptual demands are that gardens of a house should have a high visual privacy, meaning that they should be minimally exposed to visual perception from the other buildings around it. Next to this, the gardens are desired to be as large as possible in south direction.

These criteria are selected because they are major aspects concerning the task. That is, the visual privacy and size of the south garden are significant features determining the quality of life for the residents and the value of the area. Often the perceptual aspects are the most prominent factor distinguishing the quality of architectural designs, and generating the greatest added value during the lifetime of a building. This is because architecture addresses our esteem needs and needs for self-actualization essentially via visual perception properties of our environment. Therefore, perception is a matter of general relevance in architectural design. This is exemplified by the common requirement for high visual privacy.

The position of a building influences the satisfaction of both perceptual and spatial requirements, and this causes conflicts aiming to reconcile both of them at the same time. These conflicts are an exemplary situation to demonstrate the intelligent objects' ability to deal with conflicting requirements. However, requirements for perceptual properties are generally difficult to take into account, because they are vague and involve imprecision. It is problematic to compare the perceptual properties of different spatial situations in a 'fair' manner, i.e. using the same measurement metric. That is because every spatial situation contains abundant visual information that obscures the subtle differences among the situations in terms of their respective perceptual properties. This means if we assess the perceptual properties of a design by means of visual inspection only, we will have imprecise results. Therefore we are commonly not in a position to clearly weigh privacy aspects against the size of gardens for instance.

Fig. 5.1 The neighbourhood subject to design

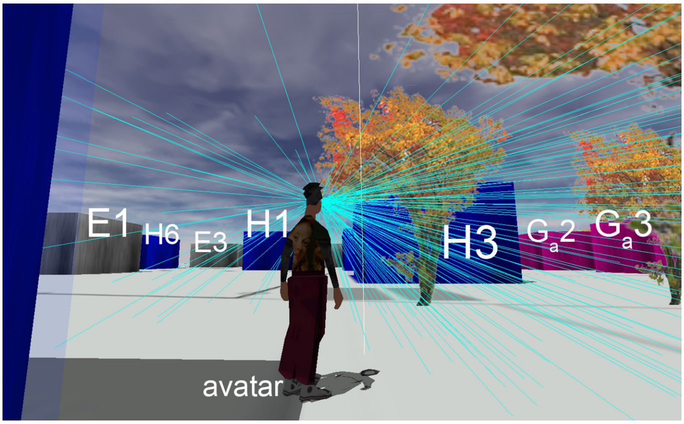


Fig. 5.2 A virtual observer is viewing the scene for perception measurement

Due to the softness of visual perception, where brain processes play an essential role, up till now perception based requirements have not been subject to computational treatment. Alleviation of this bottleneck is exemplified in the present application.

In architectural design we make use of models to predict with precision the stability and strength of our buildings, how they perform in different wind conditions, how much energy they will consume and what they will cost to build. Up till now we have no means to predict with comparable precision visual perception aspects of spaces. This is because visual perception is a phenomenon that is vaguely understood and merely verbally described in the literature [7-10]. If we wish to let perceptual requirements play a significant role in design, it is clearly necessary to treat perception in a more definitive form beyond verbal descriptions and imprecise intuitive assessments. This holds in particular when we consider trading-off perceptual qualities for other criteria, such as functionality, sustainability etc.

In the present example visual perception is treated by employing the perception theory presented in Chapter 2. Using this model an architect is able to express his/her perceptual preferences with precision. In this way subtle perceptual differences among designs can be detected. Based on this information, potential trade-offs between perception and functional requirements of different solutions can be obtained through the computational performance evaluation. This information is used in the computational design generation to reach designs that satisfy both criteria simultaneously. Based on the designs resulting from the computational process, we are able to select among the solutions with great awareness.

In the design exercise we verify how well the novel method accomplishes the design task we pose to it. This is done by comparing the performance of the designs it generates in the final generation step with a random design from the initial stage of the process. This means we observe how high the degree of satisfaction for every design requirement is, and how high the overall score resulting from the satisfaction of the elemental requirements is. As it is unknown what the highest possible solution to the problem we posed is, additional means are used to verify the performance of the method. Namely, the *convergence* of the genetic search is verified. This is accomplished by verifying if the chromosomes in the population reach a performance level that is comparable to the best solution found during the search after some time. This is a well known technique to verify the performance of a genetic algorithm.

5.1.2. Assessment of visual privacy performance

In order to obtain the visual perception properties of a design the probabilistic perception theory is implemented. The principle behind the perception model is illustrated in figure 5.2. The figure shows the same situation as figure 5.1 from the viewpoint of a virtual observer viewing the scene.

In the perception theory the visual attention an observer pays to a scene is modelled as a probability density function (pdf). This is illustrated in figure 5.2 by means of a number of vision rays that are leaving the eyes of a virtual human in random directions. The probability density of the rays is shaped in accordance with the unbiased observer model described in the theory. This means the virtual observer has no a-priori preference for objects in its field of view. The unbiased observer model is an adequate standard to use in the present application, because there is no justification for a-priori bias in the vision. The amount of vision rays reaching a certain object yields the degree of perception of this object. This is expressed by means of a probability. This probability quantifies the degree an observer is aware of the object in his/her environment. This method is implemented into the computational design process, so that the perception of a building from several viewpoints in the environment is quantified.

Figure 5.3 illustrates the computation of visual perception of the buildings *H1*, *H2*, *H3*, and *H4* from building *E1*.

Here the viewpoint of the observer is taken as the geometric centre point of the north façade of building *E1*. The curves plotted along the *z* axis are the probability density functions belonging to the perceptions of the houses *H1-H4*. They show the degree of visual attention paid to these buildings' south facades. The integral of the pdfs over the length of the south façade of each house is indicated as a shaded area. It quantifies the perception of the respective facade.

Based on the probabilistic perception in this implementation the visual privacy belonging to an area is computed. It is given by the reciprocal of the summed up perceptions of the area obtained from the relevant observation points in an environment.

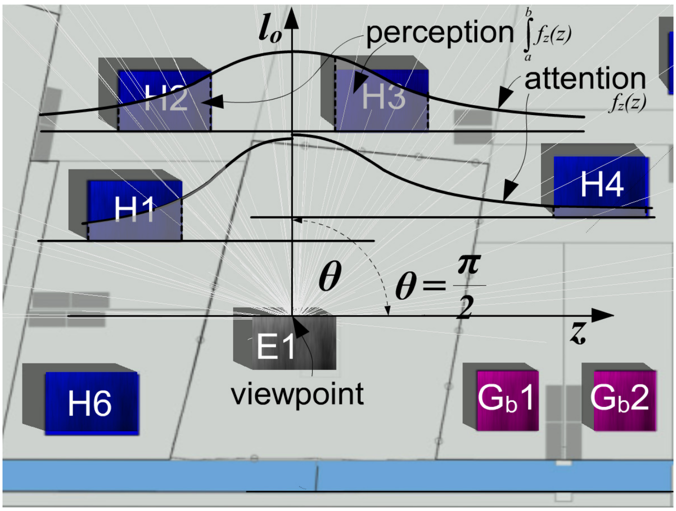


Fig. 5.3 Sketch indicating the computation of the degree of perceptions of the houses *H1, H2*, and *H3* from the viewpoint *E1*.

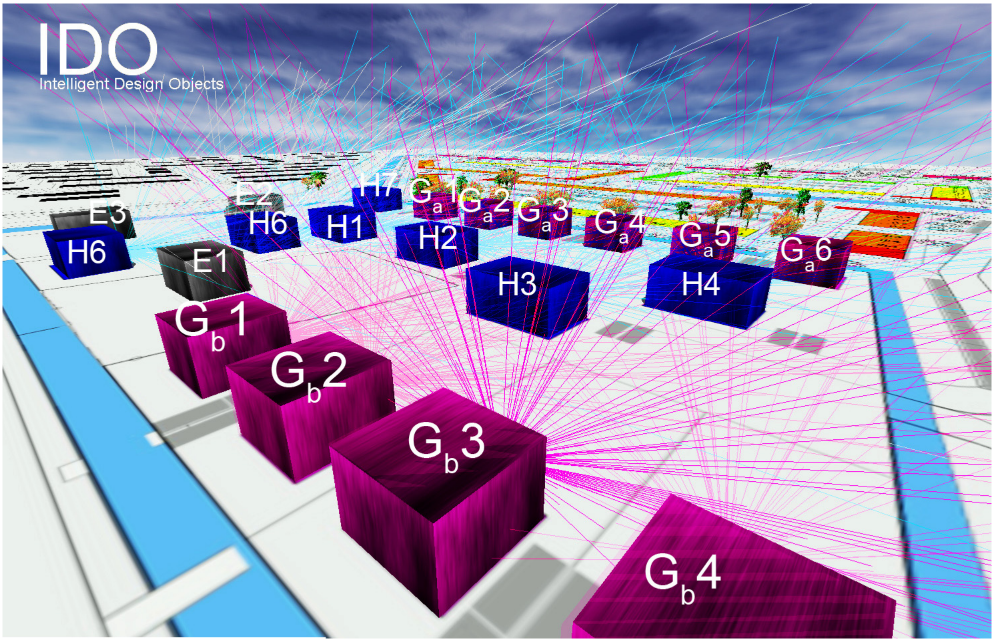


Fig. 5.4 Visual privacy computation based on the probabilistic perception model

This means the visual privacy of a façade is considered the reciprocal of the sum of attention "impinging" on the facade. In other words it quantifies how low (or high) the degree of perception of a façade is. Explicitly we calculate the visual privacy of an object *O* as

$$P_{priv}(O) = \frac{1}{\sum_{1}^{n} P(O,V_n)} \tag{5.1}$$

where $P(O,V_n)$ is the degree of perception of object *O* from the *n*-th viewpoint. In this implementation we consider the visual privacy of the south façade of the building, because in this design it is expected that living rooms and openings to the garden are oriented to the south side of the buildings. These areas are considered the most important ones with respect to privacy perception in this design. Figure 5.4 illustrates the implementation of the visual privacy computation for the houses of the housing complex.

Every south facade is perceived from several viewpoints and the visual privacy for each house is computed as given by Eq. 5.1. We note that in the computation of the perceptions in this implementation, occlusion is considered. Occlusion is detected by a simplified test of the visibility of a building from another one. The mechanism is sending a test-ray from the centre location of the first building to the viewpoint and identifying if the ray is intercepted by another building located in between them. If this is this case the perception of the building from the second one is considered to be zero.

5.1.3. Assessment of garden performance

A second aspect considered in the design of the housing complex is the size of the gardens. We consider that in general a garden located south of the building is most desirable due to exposure to direct sunlight. Therefore the performance of a garden is calculated based on

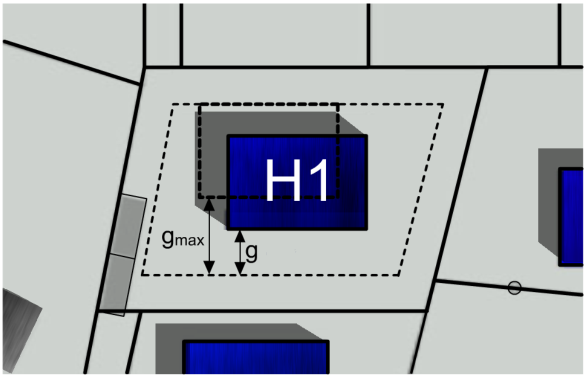


Fig. 5.5 Calculation of the garden performance

the size of the south garden. The buildings *H4* and *H5* form an exception. The lots of these houses are oriented in east-west direction. Therefore, next to the garden in south direction, the gardens west of the buildings are considered. In this case the west direction is used and not the east direction. This is done because for this design task we assume the residents of the houses *H4* and *H5* wish to have direct sunlight in their garden during the evening rather than in the morning. In order to determine the garden performance the size of the garden in south direction is normalized with respect to the maximally possible size of the garden in this direction. The maximum size of the garden in south direction is restricted by the minimum distance between the boundaries for placement in north and south direction and the width of the house. This is illustrated in figure 5.5 using house *H1* as an example. In the figure the boundary of the lot is shown as a solid line while the placement boundary is shown as a dashed line. Explicitly the garden performance *G* is given by g/g_{max}. This means the garden performance belonging to a house is calculated by dividing the extent of a garden in south direction by the maximum extent the garden can have, considering the boundary of the house's plot.

5.1.4. Establishing the model for performance evaluation

For the evaluation of the performance of alternative solutions, we consider two major aspects. A garden should have a high visual privacy, and it should have a large size in south direction. The knowledge about the performance of the design is established In the fuzzy neural model, as shown in figure 5.6.

From the figure it is seen that the performance of a neighbourhood design is determined by two major aspects. These are the performance of the gardens referring to their sizes in south direction, and the performance in terms of the visual privacy of the south facades. These aspects are located at one level below from the root node, which is the penultimate level in the tree. The performance of the overall design regarding these aspects is high if the satisfaction of the perception and garden requirements for every house are also highly satisfied. Therefore one level below the penultimate level is the terminal level. This is except with respect to the garden performance of houses *H4* and *H5*, where the garden performance has two additional sub-aspects. These aspects are the performance of the garden to the west and the south side of the house respectively. Another exception is the privacy performance of the houses G_a1-G_a6, which together form an additional sub-aspect of the privacy performance.

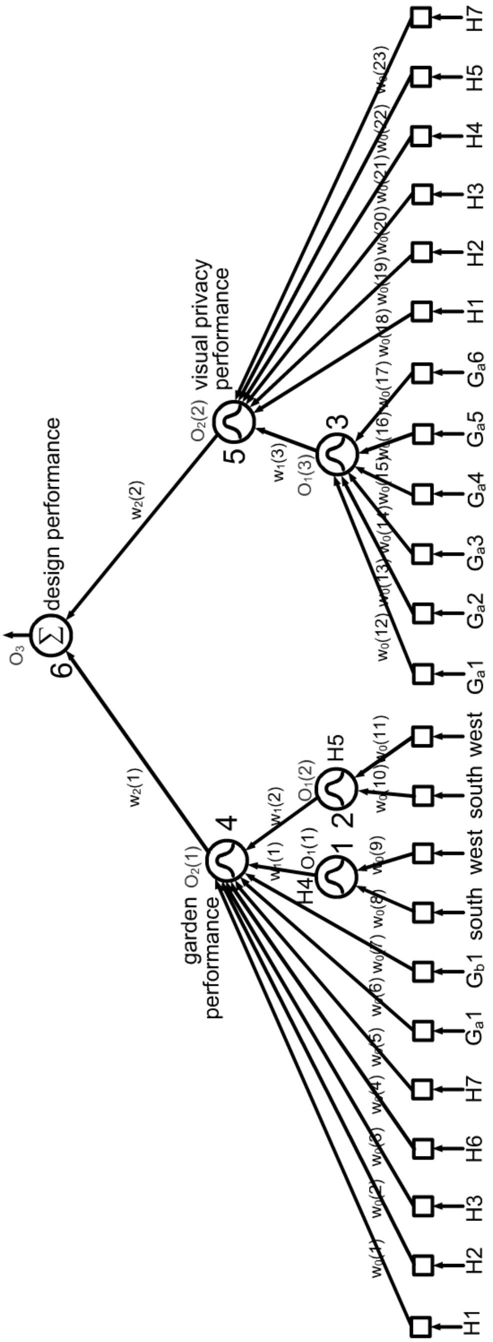


Fig. 5.6 Neural tree structure for assessment of design performance

These determinants of the design performance one level below the penultimate level are given in table 5.1.

For the tree structure established, the connection weights at each level are assessed by an architect. They are given in table 5.2. These weights indicate the relative importance of a sub-aspect compared to other sub-aspects. The structure can be considered as constitution of domain knowledge, where the connecting weights between the nodes are determined by expert judgment. This means there is no formula used to determine the weights. They are an expression of an expert knowledge representing the relative importance of the garden and privacy of a certain house for the evaluation of the design, which is due to the specific preferences of a client of the neighbourhood design, which are simulated in the present case. The houses are considered to differ with respect to their importance for privacy and garden performance in the overall design, which is a common consideration in design, due to different targeted values for a house. However, concerning the visual privacy of the group of houses G_a1-G_a6 it is noted that the designer considered every house equally significant with respect to privacy of the design, so that the weights w_{12}-w_{17} are equal.

It is to be noted that level 2 is not specified in table 5.2, since in the present case the relative importance among the objectives privacy and garden size is problematic to establish for a designer before having an impression on the nature of the conflicts among the objectives. This means it is left up to the computational cognition to hint designers what relative importance among the objects yields maximal design performance in the present task.

It should be noted that the present approach is in contrast to artificial neural networks as follows. Artificial neural networks need data and training. The neural trees used in the present approach are established from expert knowledge. That is, the structure of the tree and the weights are given based on expert knowledge. In the present design task two main aspects are important to judge the performance of the design, visual privacy and size of garden. The garden performance is high if all houses in the neighbourhood have large gardens, and the privacy is high all houses in the neighbourhood have a high privacy. Every aspect is considered in the context of design performance of the housing complex and eventually assessed between zero and one. The assessments are used as connection weights w_{ij} in the neural tree. Based on these weights a knowledge model is formed. The model should comply with the following condition: *the greater the membership value w_i of an aspect, the greater the design performance*.

Due to the peculiarity of the neural fuzzy structure, only the left half side of the

Table 5.1 Determinants of the design performance one level below the major aspects

Garden performance	**Visual privacy performance**
Garden of house *H1*	Privacy of house *H1*
Garden of house *H2*	Privacy of house *H2*
Garden of house *H3*	Privacy of house *H3*
Garden of house *H4*	Privacy of house *H4*
Garden of house *H5*	Privacy of house *H5*
Garden of house *H6*	Privacy of house *H6*
Garden of house *H7*	Privacy of house *H7*
Garden of house G_a1	Privacy of group G_a
Garden of house G_b1	

Table 5.2 Weights of the neural tree for the design performance

weight nr.	1	2	3	4	5	6	7	8	9
level 1	.11	.09	.09						
level 0	.11	.11	.11	.09	.09	.08	.15	.70	.30

weight nr.	10	11	12	13	14	15	16	17	18
level 0	.70	.30	.17	.17	.17	.17	.17	.17	.11

weight nr.	19	20	21	22	23
level 0	.14	.19	.19	.13	.09

Gaussians at the inner nodes are used during the computations. Therefore the inner nodes represent a multivariable increasing function. This ensures that greater membership value w_i of an aspect at the input to a radial basis function yields greater node output. It is noted that the model is completely knowledge-driven and highly non-linear due to the Gaussians. This is at least with respect to the non-terminal nodes and fuzzy membership functions at the terminals.

As far as non-terminal nodes are concerned the widths of the Gaussians are still to be determined. They are obtained by means of the consistency condition, which serves as boundary condition for the neural tree model.

5.1.5. Modelling the requirements at the terminal nodes

The requirements for the visual privacy are that the privacy should be as high as possible, and satisfaction becomes less when the privacy approaches to zero. These requirements are expressed by means of fuzzy membership functions at the terminal nodes.

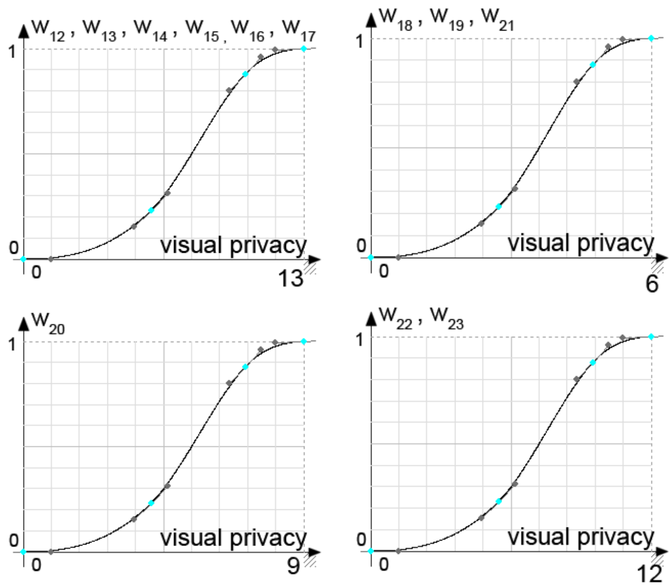


Fig. 5.7 The elemental requirements expressed by fuzzy membership functions

They are shown in figure 5.7. The shapes are selected by an architect. It is noted that they are expressions of a knowledge. This knowledge expresses under which conditions a requirement is satisfied to what extent. This means the fuzzy functions are the representations of the requirement specifications of the design. It is noted that the functions selected for the privacy performance measurement all have the same basic shape. However, the output maxima are at different locations. This is to account for different requirements that are due to the different housing types and lot conditions involved. For example the membership function for the leaf node outputs w_{18}, w_{19}, and w_{21} expresses that visual privacy is required to be value 6 or greater as an ideal situation. Then the leaf node output is one, representing maximum satisfaction of the corresponding requirement. If the privacy is less than 6 the output of the respective leaf node is less than one, as specified by the membership function. For the other houses more privacy is required as can be seen from the location of the respective maximum.

Concerning the garden performance the fuzzy membership function used is trivial; namely $w_i=x$ because the garden performance $G=g/g_{max}$ is already normalized between zero and one. Therefore G directly serves as the node output of the respective terminal node.

5.1.6. Generating Pareto optimal designs

Before the neural tree can be used for performance evaluation the consistency condition has to be imposed on the neural tree. This can be carried out by adaptive or genetic learning. In the present implementation adaptive learning is used for high accuracy. As result of the learning process, the width of each individual Gaussian at each non-terminal node is established. In this way, the feed-forward fuzzy logic operations are clearly defined exhibiting features of transparency in the model. The approximation error for this data set is relatively higher for the extremely low and high input/output pairs. This is seen from table 5.3. The resulting widths of the Gaussians at the non-terminal nodes are given in table 5.4.

Having established the fuzzy neural tree, the design task is to maximize the output at the root node by identifying optimal locations of the buildings. This is accomplished by genetic search. The boundary of the space for the locations of the houses is given in table 5.5. This is given by the minimal and maximal x and z coordinates for the positions of the houses $H1$-$H7$, G_a1-G_a6 and G_b1-G_b4. The boundaries are selected, so that the facades of the buildings are at a distance greater than 3m from the boundary of the lot.

Table 5.3 Adaptive learning results from the imposition of the consistency condition

Given for all inputs and the root output	**Approximation**	**Error**
$1.00\ 10^{-1}$	$1.68\ 10^{-1}$	$-6.82\ 10^{-2}$
$2.00\ 10^{-1}$	$2.30\ 10^{-1}$	$-3.00\ 10^{-2}$
$3.00\ 10^{-1}$	$3.06\ 10^{-1}$	$-5.96\ 10^{-3}$
$4.00\ 10^{-1}$	$3.95\ 10^{-1}$	$5.47\ 10^{-3}$
$5.00\ 10^{-1}$	$4.91\ 10^{-1}$	$8.62\ 10^{-3}$
$6.00\ 10^{-1}$	$5.89\ 10^{-1}$	$1.12\ 10^{-2}$
$7.00\ 10^{-1}$	$6.79\ 10^{-1}$	$2.11\ 10^{-2}$
$8.00\ 10^{-1}$	$7.80\ 10^{-1}$	$1.98\ 10^{-2}$
$9.00\ 10^{-1}$	$9.43\ 10^{-1}$	$-4.34\ 10^{-2}$

Table 5.4 Resulting widths of the Gaussians at the non-terminal nodes

Node nr.	**1**	**2**	**3**
σ	$7.22\ 10^{-2}$	$3.36\ 10^{-1}$	$1.95\ 10^{-1}$

Node nr.	**4**	**5**	
σ	$1.77\ 10^{-1}$	$1.55\ 10^{-1}$	

This is according to legal regulations applying to this design case.

For the sake of simplicity of the implementation the boundaries of the placement are taken parallel to the *x* and *z* axis. The *z* axis is in north direction, and the *x* axis is in east direction. The output at the root node, which expresses the design performance by a scalar number, can be used as the representation of the fitness of the respective chromosome. In this way the genetic algorithm (GA) uses the knowledge embedded in the neural tree during its search for high performance solutions, while the search is essentially treated as a single-objective problem. However, a GA used for single-objective optimization is sensitive to small changes in the objective function coefficients, which in the present case refers to the weight factors of the connections in the neural tree. This means we should be able to specify accurately in advance how important the privacy is compared to the size of the garden. This is not feasible, since we are unable to foresee what implications such a commitment will have on the solution we will obtain. Therefore it is appropriate to apply a multi-objective approach. This way the GA does not only converge to a single solution, but it provides information about alternative solutions that are equally valid in Pareto sense. The latter feature is relevant, so that we remain flexible in our decision making to take a final decision with great awareness. In our case garden performance and visual privacy performance are to be maximized simultaneously. This means we are interested in a variety of alternative solutions that are equal in Pareto sense.

In the multi-objective implementation the outputs of the nodes 4 and 5 of the neural tree are subject to maximization. Their values are used in the fitness determination procedure of the genetic algorithm. Employing the fuzzy neural tree in this way the genetic search is equipped with human-like reasoning capabilities during its search, so that the

Table 5.5 Solution space

House	**H1**		**H2**		**H3**	
	x	**z**	**x**	**z**	**x**	**z**
min	25.0	26.0	26.0	46.0	56.0	47.0
max	31.0	34.0	36.0	56.0	69.0	56.0

House	**H4**		**H5**		**H6**	
	x	**z**	**x**	**z**	**x**	**z**
min	81.0	34.0	86.0	52.0	21.0	6.0
max	117.0	38.0	114.0	57.0	28.0	16.0

House	**H7**		**G_a1**		**G_b1**	
	x	**z**	**x**	**z**	**x**	**z**
min	3.0	70.0	27.0	67.0	76.0	7.0
max	10.0	80.0	32.0	81.0	79.0	22.0

behaviour resulting from the search is deemed an intelligent behaviour. The two objectives, to simultaneously maximize the garden and the privacy performance of the design, are conflicting. The conflict is that satisfaction of one objective diminishes satisfaction of the other one. For example if a certain house moves north on its plot increasing the size of its south garden, at the same time it approaches closer to its neighbouring buildings in north direction. This reduces the neighbouring buildings' degree of visual privacy.

We use a multi-objective genetic algorithm to identify the Pareto optimal frontier. This is accomplished by assigning the fitness to a chromosome in a population depending on how many other chromosomes are dominating it. Ranking by Pareto dominance on problems with an increased number of objectives might not longer be effective [11]. One of the important issues to address in this respect is the diversity of the Pareto solutions with minimal aggregation at the Pareto front [12].

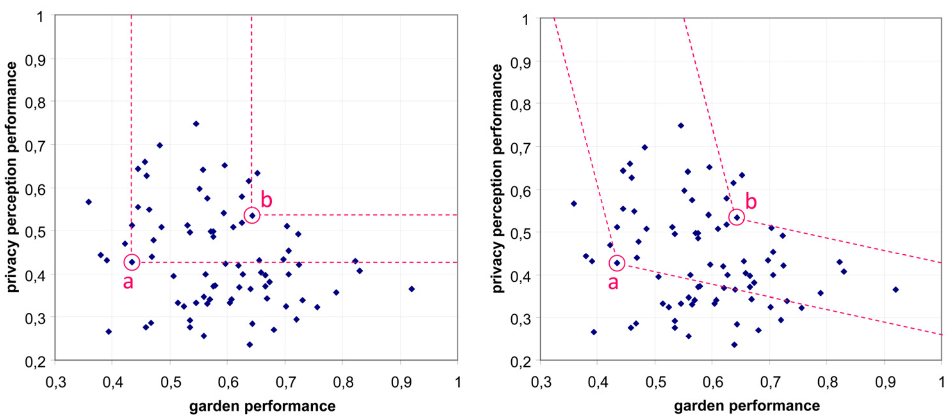
5.1.7. Application of greedy versus relaxed dominance

In Pareto ranking dominance is conventionally measured with an angle of tolerance of $\theta=90°$. This is shown in figure 5.8a for our two-dimensional optimization problem.

Figure 5.8a shows the garden and privacy performance of the initial population in the search. The dashed lines indicate the region, in which dominant solutions are counted for the respective chromosome. Figure 5.8b shows the same situation for $\theta \gg 90°$. In greedy dominance, i.e. for angle of tolerance $\theta=90°$, solution *a* is dominated by 33 other solutions, whereas chromosome *b* is dominated by 1 solution. The novel approach in this work is to widen the scope of the tolerance, within which the degree of dominance is determined. This is shown in figure 5.8b. In this figure chromosome *a* is dominated by 46 solutions and chromosome *b* is dominated by 2 solutions. The amount of dominating chromosomes denoted by *n* is converted to fitness *F* by

$$\boldsymbol{R}_{fit} = \frac{1}{N(\theta)+n} \tag{5.2}$$

N(θ) is a parameter that should be carefully chosen with respect to the angle of tolerance. This is because it scales the relative importance of the amount of domination. The fitness is



(a) (b)

Fig. 5.8 Privacy performance and garden performance of the population in the first generation of the GA, where the angle of tolerance θ=90° (a); and θ>>90° (b) thereafter converted to a probability for reproduction applying the well-known roulette wheel selection principle [13]. The roulette wheel selection is used as a heuristic method, which means it is a straight forward implementation conforming to the genetic process applied to a living population as a whole. Because of that it may take more time than other selection strategies, such as tournament selection for instance. However, since the convergence is very sensitive to the genetic parameters to make comparison between the strategies is difficult issue in general. This is particularly so regarding comparison of take-over time [14]. Moreover, comparison is also application dependent, which affects the takeover time, and the selection process is not the only player at the effectiveness of the convergence. The discussion in the reference is for single objective case. However, for multi-objective case, comparison becomes much more involved, and the concept of take over time becomes problematical.

In the relaxed dominance approach, if we design the *F* matrix in Eq. (4.7) in Chapter 4 as

$$\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} 1 & -.176 \\ -.176 & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \tag{5.3}$$

this implies an angle of tolerance of 10^o between the respective orthogonal and non-orthogonal system axes is employed. This way the cosine direction matrix given by Eq. (4.14) becomes

$$D = \begin{bmatrix} .984 & -.173 \\ -.173 & .984 \end{bmatrix} = Q^T \tag{5.4}$$

The performance score of a solution is computed by a defuzzification process given by

$$w_1(1) f_1 + w_2(2) f_2 = p \tag{5.5}$$

where $w_1+w_2=1$; f_1 and f_2 are garden performance and privacy performances respectively; p indicates the design performance. In our example we take w_1=0.6 for the privacy, and w_2=0.4 for the garden performance.

The corresponding weighted objectives F_1 and F_2 are given by

$$\begin{bmatrix} f_1^P \\ f_2^P \end{bmatrix} = \begin{bmatrix} .984 & -.173 \\ -.173 & .984 \end{bmatrix} \begin{bmatrix} F_1^P \\ F_2^P \end{bmatrix}, \tag{5.6}$$

and the inverse Matrix operation becomes

$$\begin{bmatrix} F_1^P \\ F_2^P \end{bmatrix} = \begin{bmatrix} 1.049 & .184 \\ .184 & 1.049 \end{bmatrix} \begin{bmatrix} f_1^P \\ f_2^P \end{bmatrix}. \tag{5.7}$$

The results given by Eq. (5.7) are interesting because the Pareto front determination in the evolutionary multi-objective optimization is considerably simplified. Eq. (5.7) is the coordinate transform used to verify dominance among solutions using the relaxed approach.

The performance of the search process is indicated in figure 5.9, where the best fitness that occurred during the search is plotted together with the average fitness of the chromosomes for each generation. From the figure the convergence of the genetic search is seen.

With reference to Eq. (5.2) the parameter $N(\theta)$ in both cases, namely for an angle of tolerance $\theta=90^o$ and $\theta=110^o$, is taken to be constant as $N=4$. Changing the value of N with respect to the angle θ the performance of the Pareto search can also be tuned, which is not considered in this research. However one should note that the selection of θ itself affects

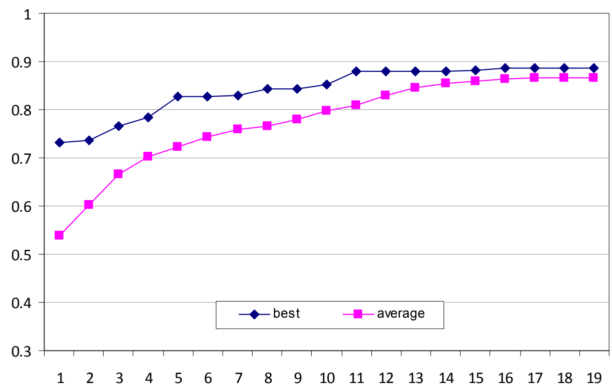


Fig. 5.9 Performance of the solution with the highest design performance for every generation

the performance of the Pareto search. With this in mind, to determine $N(\theta)$ needs attention to consider its effect concertedly.

After 20 generations the GA finds a convex Pareto optimal frontier. This is shown in figure 5.10a and 5.10b. In figure 5.10a the solutions are more scattered compared to those in figure 5.10b. This is due to the greedy dominance of the search process for angle of tolerance θ=90°, i.e. there is no relaxation in the search, as the angle of relaxation ϕ is zero. In case of relaxation, the angle of relaxation is taken as $\phi=10^{o.}$ The figures are combined and shown in figure 5.11 for comparison.

From figure 5.11 the positive effect of relaxation is clearly seen providing smooth and non-aggregated solutions. In the greedy case solutions are scattered. It is also noted that in the relaxed case the Pareto front is located further into the non-dominated region in the objective space, meaning these solutions have a higher performance score. In the figure several design performance scores belonging to respective design solutions are indicated. Performance score is defined in this work as the weighted summation of the garden and privacy performance and named as *design performance*. This score is the output value at the root node O3 of the neural tree. It is especially noted that the performance score given in figure 5.11 next to a certain solution is the maximal design performance value the

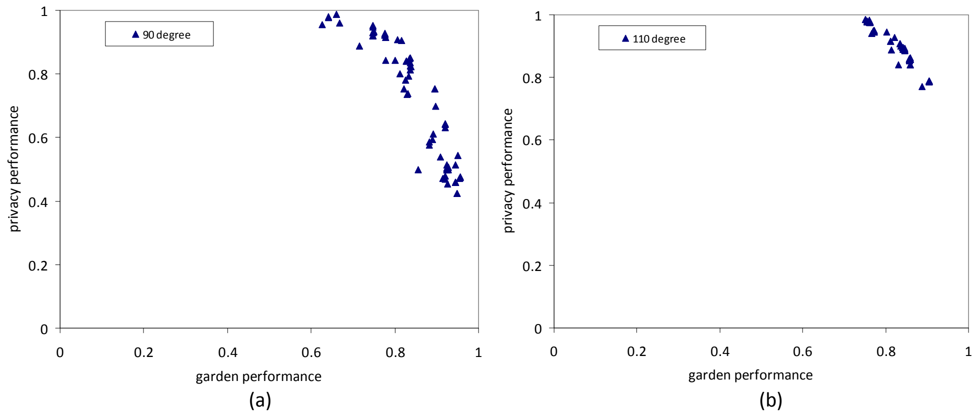


Fig. 5.10 Garden and privacy performance of the solutions after 20 generations; angle of tolerance θ=90° (a); θ=110° (b)

solution can have for any possible weight configuration at the penultimate level. This value is obtained using

$$P_{max} = \frac{O_2(1)^2 + O_2(2)^2}{O_2(1) + O_2(1)} \cdot \qquad (5.8)$$

From the neural tree the performance score is computed by

$$O_3 = w_2(1)O_2(1) + w_2(2)O_2(2) \qquad (5.9)$$

If a certain performance score would be prescribed, O_3 becomes a constant, so that eq. 5.9 can be written as

$$w_2(1)O_2(1) + w_2(2)O_2(2) = c\,, \qquad (5.10)$$

which represents a line with respect to the outputs at one level below the root node; namely *$O_1(1)$* and *$O_2(2)$*.

In the following analysis c is selected to be c=0.83. The line represented by Eq. 5.10 is shown in figure 5.12 for various weight compositions. As an example it is considered privacy constitutes 60% of the design performance, so that *$w_2(1)$=0.6*, whereas the privacy is considered 40% of design performance, so that *$w_1(1)$=0.4.* With this weight composition the intercept is 1.38 as indicated in the figure. The corresponding solution is the intersection of Pareto surface and the line given by Eq. (5.9). The intersection is denoted by S_1 in the figure. For another weight composition the

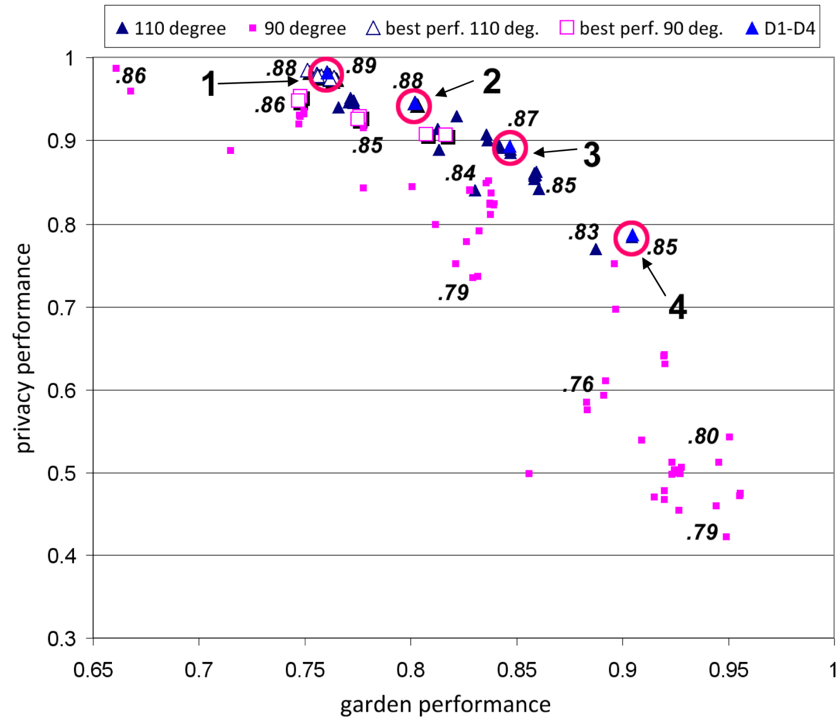


Fig. 5.11 Comparison of the Pareto optimal frontiers for the angle of tolerance θ=90° vs. θ=110° after 20 generations; a number of design performance scores are indicated

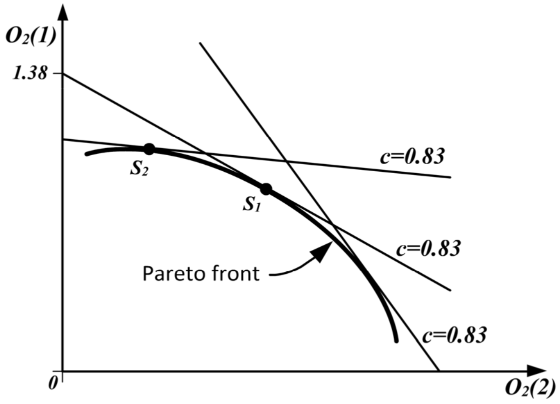


Fig. 5.12 Particular solutions on the Pareto front for different weight compositions, and corresponding to a fixed performance score, which is .83 in this case

intersection point is shown as S_2 in the figure. These points corroborate with the results shown in figure 5.11.

The scores of the total number of solutions are shown in figures 5.13. For comparison the figure shows the scores on the Pareto front, i.e. for the 80 chromosomes, in the relaxed case as well as the greedy case. We note that the performance scores are generally higher in the relaxed versus the greedy case. Also in the relaxed case the scores fluctuate in a narrower domain compared to the greedy case. This means for the relaxed case the solutions along the Pareto surfaces are all approximately equally valid. This means they provide flexibility in the selection of a solution during design. The scores of the random solutions in the initial population of the search are also shown in the same figure, putting the accomplishment of the evolutionary search into perspective.

One should note that the performance score at the root node of the tree does not play a role in the genetic multi-objective optimization.

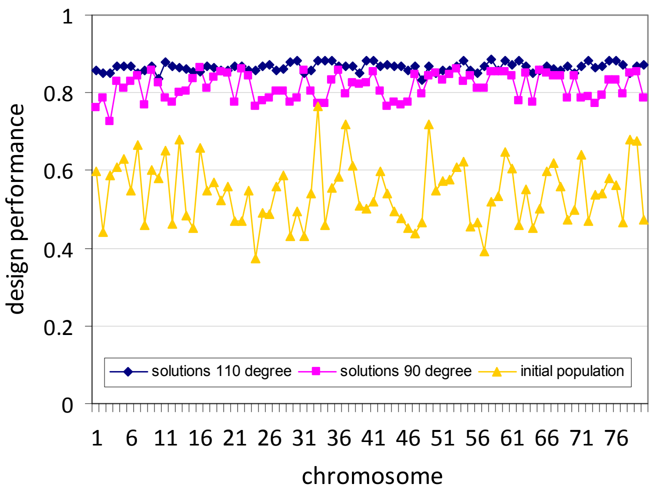


Fig. 5.13 Design performance scores for the solutions in the Pareto front found using an angle of tolerance θ=110° vs. θ=90°

However, the Pareto front offers a number of design options with fair performance. This leaves the final choice dependent on other environmental preferences. In this respect it is emphasized that we are not able to tell in advance how important each of the criteria should be taken. This is due to the vagueness of the criteria, visual privacy and garden performance. However, as soon as we become aware of the implications of these criteria on the solutions, we are clearly in a much better position to take a decision about the priority among criteria. That is, when we compare the Pareto optimal solutions, we no longer have to bear in mind the relation between de parameters of the design and the performance aspects of the design, as this relation is 'embodied' in the form of the Pareto optimal solutions already. This means we are able to pay more attention to the relative importance among the criteria, selecting a solution that suits our preferences with great awareness.

5.1.8. Resulting Pareto-optimal designs

Figure 5.14 shows a design at the beginning of the genetic search process, which is a random configuration. It has a design performance of 0.41, which is the maximal performance output value at the root node of the tree. One notes that the marked areas on the lot are the locations originally proposed by the urban design office as solutions.

Four Pareto optimal designs are shown in figures 5.15-5.18. In the figures we are looking in north direction. The designs belong to different regions on the Pareto front shown in figure 5.11. The solution indicated by number 1 in figure 5.11 corresponds to figure 5.15; the one indicated by number 2 belongs to figure 5.16. The solution indicated by number 3 corresponds to figure 5.17; the one indicated by number 4 belongs to figure 5.18. It is noted that next to the design proposed by the intelligent objects in the figures another design is indicated as well. This second design is indicated by rectangles on the respective lots on the ground plane. It is a proposition by the design professionals using conventional means. It is shown on every figure for comparison.

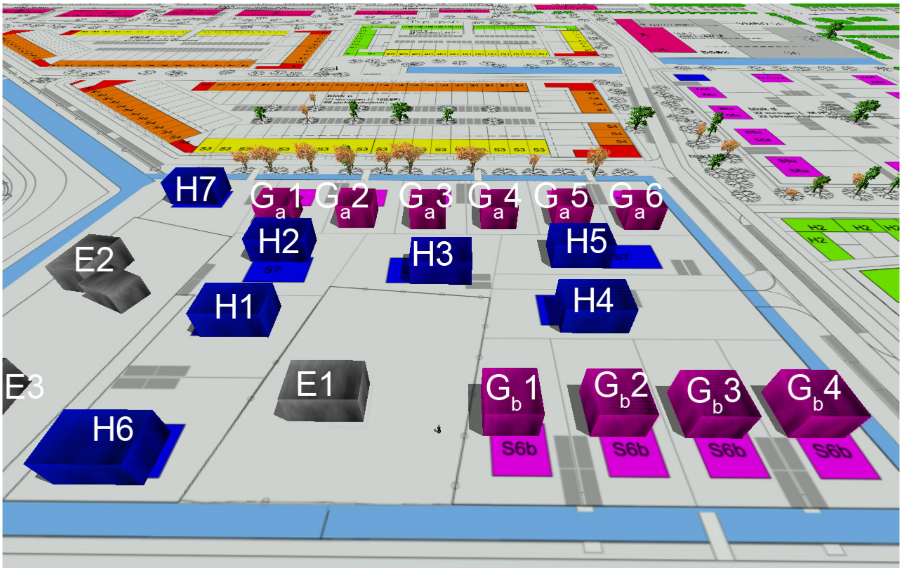


Fig. 5.14 A design at the beginning of the Genetic search process with a performance of 0.41

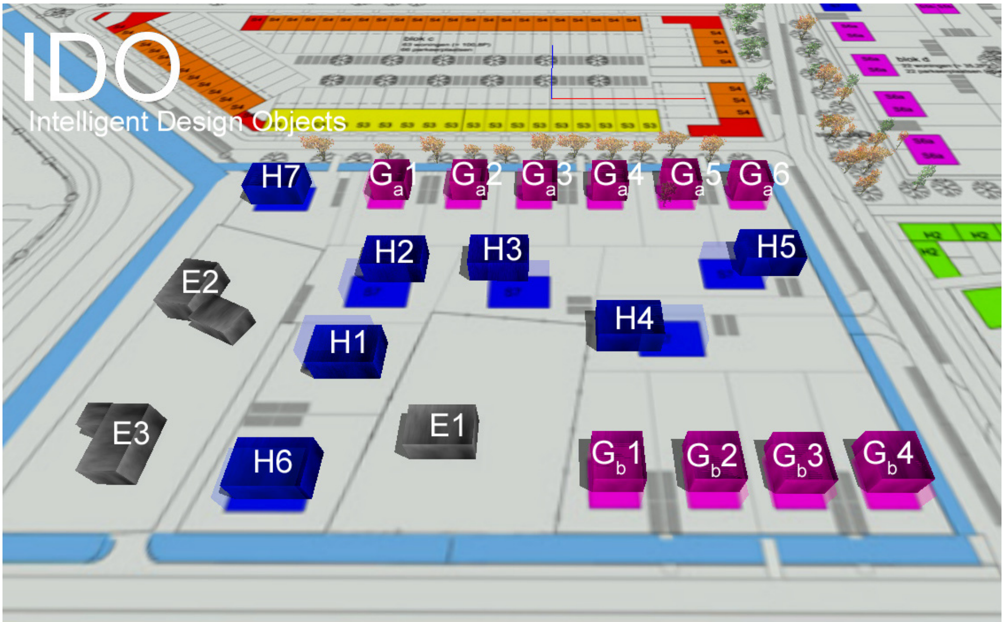


Fig. 5.15 Pareto-optimal design nr. 1

Design nr. 4 has the highest garden performance of the four designs shown. This is because houses have large south gardens, respectively west gardens in the case of house *H5*. However, as trade-off in this design the visual privacy is relatively low compared to the other designs.

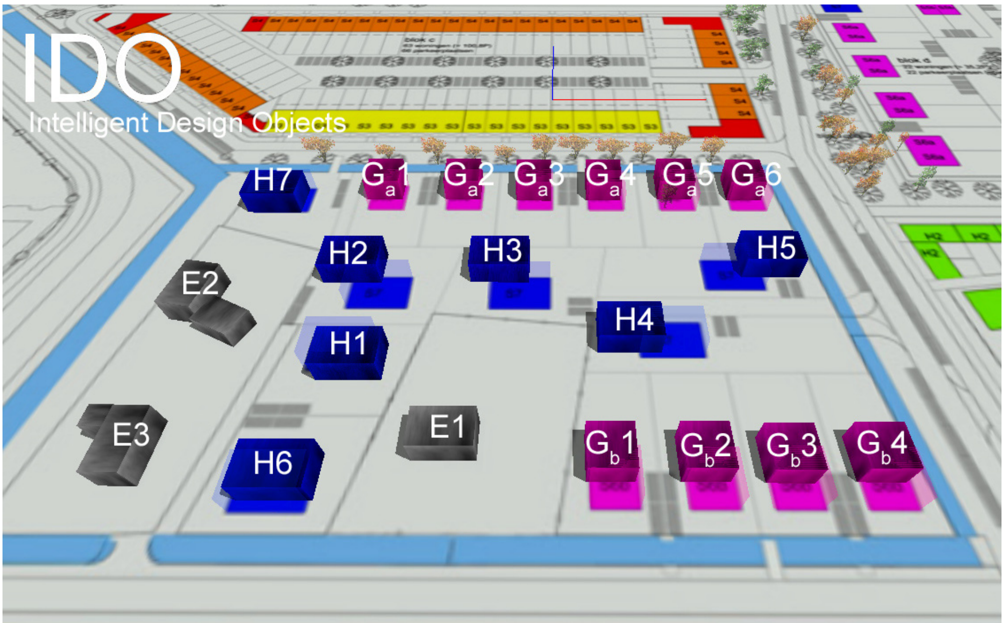


Fig. 5.16 Pareto-optimal design nr. 2

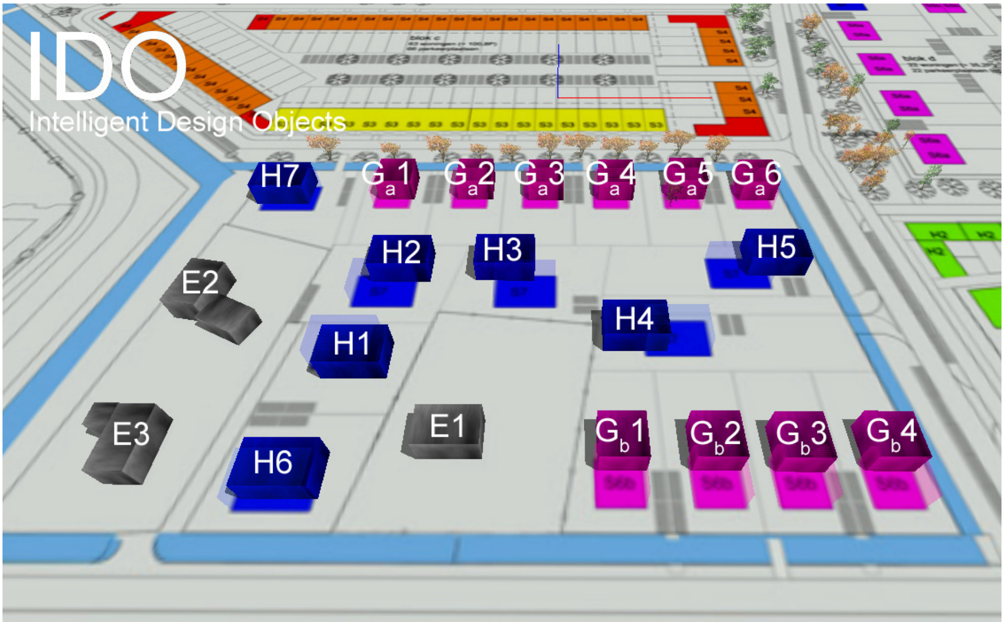


Fig. 5.17 Pareto-optimal design nr. 3

This is because the houses *H6* and G_b1-G_b4 are located quite close to the south façade of neighbouring buildings reducing their privacy.

Design nr. 3 has a higher visual privacy compared to nr. 4, while its garden performance is reduced. This is due to the more southern position of *H6* in design nr. 3. This reduces the south garden of H6 and increases the privacy of *H1*.

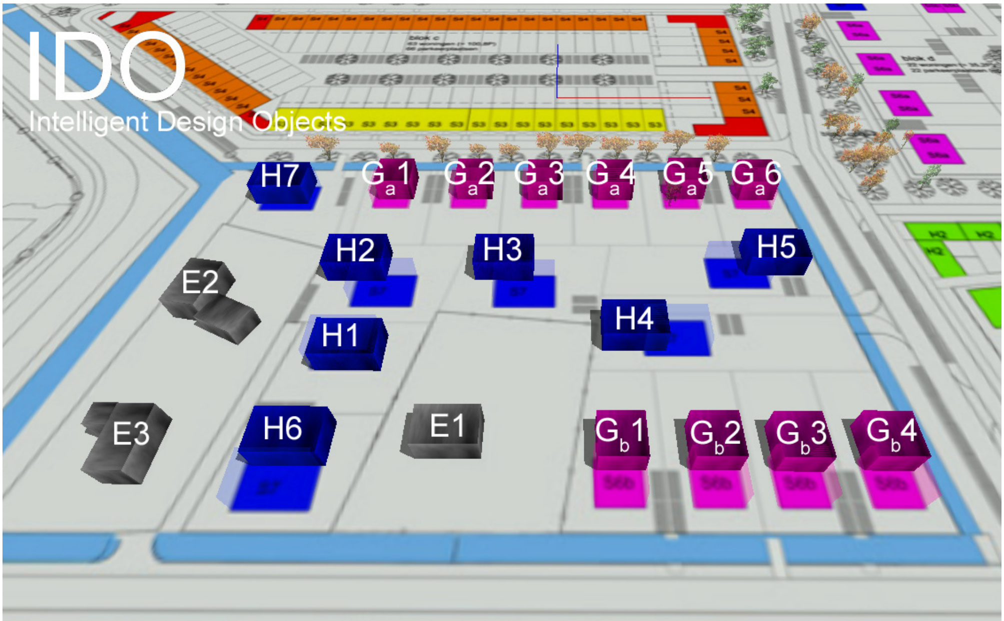


Fig. 5.18 Pareto-optimal design nr. 4

Also *H2* and *H3* are positioned in such a way that they increase the privacy of houses *H7*, as well as G_a1-G_a6.

Design nr. 2 is similar to nr. 3 with the difference that houses G_b1-G_b4 are located further south compared to nr.3. This clearly reduces the garden performance in nr. 2 yielding increased privacy in return. Although the difference in location is only a few meters, the gain in privacy is notable since the perception caused by G_b1-G_b4 affects many houses on the scene.

Design nr. 1 is similar to design nr. 2. The difference is that the houses G_b1-G_b4 are located even further south in nr. 1, so that for this design the privacy is further increased at the cost of reduced garden performance. Another difference is that H2 is located in such a way that its own privacy as well as the privacy of *H7* and G_a1-G_a6 are higher than in design nr. 2.

We note that the designs differ essentially regarding the positions of houses H6, H2, and G_b1-G_b4. This dependence, and the fact that there is a unique optimal location for the other buildings despite the multi-objective nature of the task is difficult to foresee prior to executing the computational design process. This is because modifying a single house on its plot influences the visual privacy of several other houses, and the combined effect of modifying 17 houses at the same time is difficult to consider. In particular it is formidably difficult to keep in mind all requirements and their associated tolerance for partial satisfaction. Due to the established Pareto front we are able to identify common features among the Pareto solutions. This is referred to as *innovization* in the literature [15]. Innovization is a desirable feature of Pareto optimality, since it allows gaining deeper insight into the design compared to conventional design approaches, allowing to simplify a design problem and to identify the sensitive parameters for solutions in a design problem.

Having a repertoire of Pareto optimal solutions, we are in a position to select among them. We may select with the certainty that each of them is Pareto-optimal with respect to the design criteria we put forward. In order to make a decision about which design to pick, higher-level design criteria are brought into play. In the present case we consider the relative importance of privacy and garden performance. Based on this higher-level criterion we select a certain design located on the Pareto front, which is both non-dominated and suits our preferences regarding the importance of privacy versus garden extent. This means the design selected suits both the design criteria modelled and our second-order preferences.

In order to efficiently identify a most appropriate solution among the Pareto optimal ones it is desirable to have a suitable starting place for this investigation. Therefore we first consider solution nr. 1. This is because this very solution has the highest maximal performance among the Pareto solutions according to Eq. (5.8). This means in our present design task, the solution nr. 1 is the most prominent starting point for our exploration. The corresponding weights for the connections between penultimate nodes and root node are obtained from normalized outputs at the penultimate level of the tree. Namely .98/(.98+.76)=.56 for privacy and .76/(.98+.76)=.44 for the garden aspect. This means computational cognition proposes to give somewhat higher gravity to privacy than to the garden size if we are looking for a solution with maximal performance among the Pareto solutions, which Is generally the case in design. However, for any reason we may find the gravity of privacy too much for our indented purpose. Then we may modify the priority vector to take another ratio, for example 50/50. While initially our favourite solution was

Table 5.6 Comparison of the initial design and Pareto optimal design nr. 1

Node	Output	Initial design	Pareto-optimal design
Design performance	O_3	.41	0.89
Garden performance	$O_2(1)$	.46	0.76
Visual privacy perf.	$O_2(2)$	.33	0.98
Garden H1	w_1	.92	0.73
Garden H2	w_2	.90	0.91
Garden H3	w_3	.10	0.93
Garden H6	w_4	.05	0.35
Garden H7	w_5	.80	0.88
Garden G_a1	w_6	.05	0.95
Garden G_b1	w_7	.05	0.25
South garden H4	w_8	.34	0.98
West garden H4	w_9	.19	0.03
South garden H5	w_{10}	.30	0.98
West garden H5	w_{11}	.27	0.96
Visual privacy G_a1	w_{12}	.01	0.92
Visual privacy G_a2	w_{13}	.20	0.50
Visual privacy G_a3	w_{14}	.12	0.71
Visual privacy G_a4	w_{15}	.18	1.00
Visual privacy G_a5	w_{16}	.01	0.97
Visual privacy G_a6	w_7	.69	0.77
Visual privacy H1	w_{18}	.99	0.91
Visual privacy H2	w_{19}	.78	0.95
Visual privacy H3	w_{20}	.70	0.91
Visual privacy H4	w_{21}	.28	0.97
Visual privacy H5	w_{22}	.06	0.95
Visual privacy H7	w_{23}	.61	1.00

nr. 1, according to our new fixed priority vector another design on the front may become our favourite one.

It is noted that changing the priority between privacy and the garden performance and obtaining another solution is a unique form of human machine interaction. Through this act we are asking the machine to provide a solution that has a larger garden compared to the solution initially proposed by machine cognition, while we wish to sacrifice a little as possible of visual privacy. Such higher-level trade-off considerations are conventionally reserved for human intelligence. This is because the performance concept is soft involving linguistic variables and a large amount of non-linear relations between the concepts and the design variables.

The results demonstrate that the intelligent system identifies Pareto optimal designs and proposes a design with maximal design performance among them without having information on the relative importance among criteria. It is emphasized that the approach remains transparent, meaning it is clear why a certain solution has a higher/lower performance than another one. The transparency is seen from table 5.6, where the outputs of the tree nodes for the initial design and design nr. 2 are given as an example. From the

Table 5.7 The parameters of design nr. 1

House	H1		H2		H3	
	x	z	x	z	x	z
value	30.6	31.8	35.4	55.1	56.8	55.4

House	H4		H5		H6	
	x	z	x	z	x	z
value	82.1	37.9	112.7	56.9	23.9	9.5

House	H7		G_a1		G_b1	
	x	z	x	z	x	z
value	4.2	78.8	30.3	80.4	78.1	10.8

table we also note that the intelligent system increases the performance of the final design compared to the initial design with respect to almost all requirements. Due to this transparency we are able to modify the design criteria with great awareness. This is significant if we are, for any reason, not satisfied with the Pareto optimal solutions found in the search.

Having explored a Pareto front of solutions we have an *overview* of the trade-off potentials inherent to the task. This way we are able to effectively modify the requirements, so that in a following execution of the system we are more likely to obtain satisfactory solutions.

It is interesting to compare the selected Pareto optimal design with the design obtained by conventional means. First of all it is interesting to note that the computational design with maximal performance, namely nr. 1, is quite similar to the conventional one proposed by the human architects. This is particularly apparent for the groups of houses G_a1-G_a6 and G_b1-G_b4. The similarity can be interpreted in the way that the criteria used coincide with the ones used by the architects. Next to that the second-order preferences also appear to coincide. In particular, apparently the designers favoured visual privacy over the south gardens, as this is seen from the positioning of the southern group of houses G_b1-G_b4. The differences between the designs concern houses *H2, H3, H4,* and *H5.* The locations of these houses are located more extremely at the edge of the plots maximizing the garden size for the houses. This way the gardens of these houses are larger in the computational case.

One may ask what the added value is to use computation versus conventional means in case the latter arrives at a similar result. There are two major aspects. One is that in the computational case we have an increased certainty that our requirements are fulfilled. This is challenging to obtain in the conventional case, since we are hardly able to verify the satisfaction of all requirements at the same time. Since the reasoning leading to a design solution in the conventional case is usually vague, a second benefit of the computational approach is that we remain flexible when we obtained the result. That is, we are able to consider equivalently valid alternatives, exercising second order preferences with the certainty that our elemental requirements are duly taken into account already. This means we can take a final decision with greater confidence having considered implications of alternative preferences. Also due to the results from computational cognition we are able to compare our favourite solution in perspective to a maximally attainable performance. Finally, when we arrived at the Pareto solution, we are able to integrate new requirements based on the previous ones. Conventionally, if we change our mind regarding the requirements, we would have to restart our design from scratch. This means we should remember all our requirements. This is a formidable task due to the elusive character of design requirements. Therefore it is difficult to ensure our previous reasoning is still adequately reflected. In the computational case, if we add a requirement, due to the local properties of the Gaussian at the inner tree nodes, we are able to expand the requirement model without the need to remember all other requirements during this operation.

It is emphasized that in the cognitive approach we are making use of the intelligent computational processing that takes all our criteria into account. Therefore we can take final decisions with great awareness compared to the conventional case. It is also emphasized that the cooperation between an architect and the intelligent objects as demonstrated is a cognitive act, because second-order preferences are brought into play, while the primary objectives are satisfied at the same time.

5.2. Application 2

5.2.1. Design task

This design task concerns the design of an interior space. The space is based on the main hall of the World Trade Centre in Rotterdam in the Netherlands. The exterior of the building is shown in figure 5.19.

Figure 5.20 shows the main entrance hall of the building seen from its entry. The core of the high-rise tower visible in figure 5.19 is also seen in figure 5.20, where it is labelled *building core*. In the figure a virtual human observer can be seen, who is viewing the interior space of the entrance hall. The perception of the virtual observer plays a role in the design. This perception plays a role considering certain perception related requirements. An example is that we demand that the stairs are not very noticeable from the entrance of the space. In figure 5.20 a number of sightlines show the visual perception computation involved that is based on the perception model described in chapter 2.

Fig. 5.19 Exterior of the World Trade Centre of Rotterdam

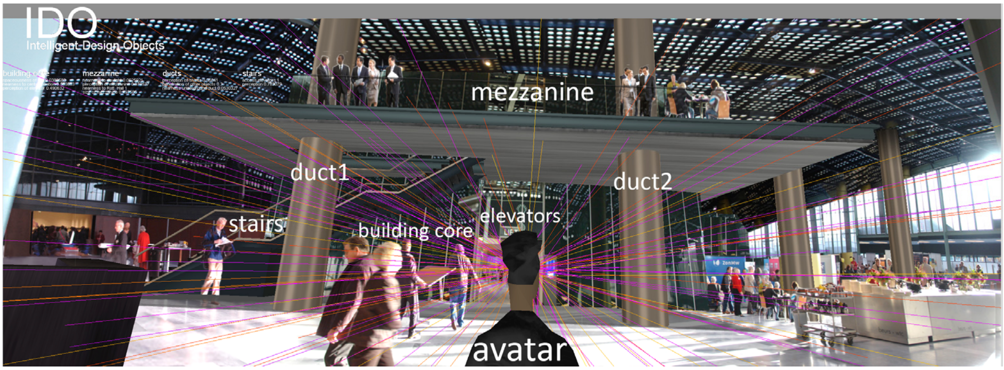


Fig. 5.20 The design objects of the task

In the present design task the boundary of the hall already exists. The task concerns the optimal placement of a number of design objects in the space so that a number of function and perception related criteria are satisfied. These objects are as follows.

(i) A building core, hosting elevator-access to a high-rise building. The high-rise tower penetrates the entrance hall, so that the hall functions as both access to the world trade centres' exhibition spaces on the ground floor, as well as access to the office tower.
(ii) A mezzanine, where lunches or informal gatherings can be organized.
(iii) Vertical ducts serving for ventilation of the underground parking space.
(iv) Stairs leading from the entrance hall to the mezzanine.

The object parameters subject to optimal identification in the present design task are given in table 5.8 per object.

Table 5.8 Parameters subject to genetic identification

building core	mezzanine	ducts	stairs
position	position	position	position
		material	

5.2.2. Establishing the model for performance evaluation

In this design example the aim is to position the objects, so that they fulfil a number of requirements. The requirements for example concern the spaciousness of the entrance hall, which is desired to be high. This is to let people experience a generous space when entering the building. The degree of perception of the stairs is desired to be low in the present example. The stairs should not be very much noticeable, because in the present case the mezzanine is considered a secondary floor, which should be visited only by people who have a specific reason to go there. It is not meant to be frequented by the general public. Another requirement is that we desire the mezzanine be large enough in size, and that its location is nearby an adjacent space.

In the fuzzy neural model, such requirements are represented as follows. The neural tree structure for this case is established as shown in figure 5.21. It is emphasized that the neural tree modelling process is not subject to automation. The weights and the neural tree structure are knowledge from a designer.

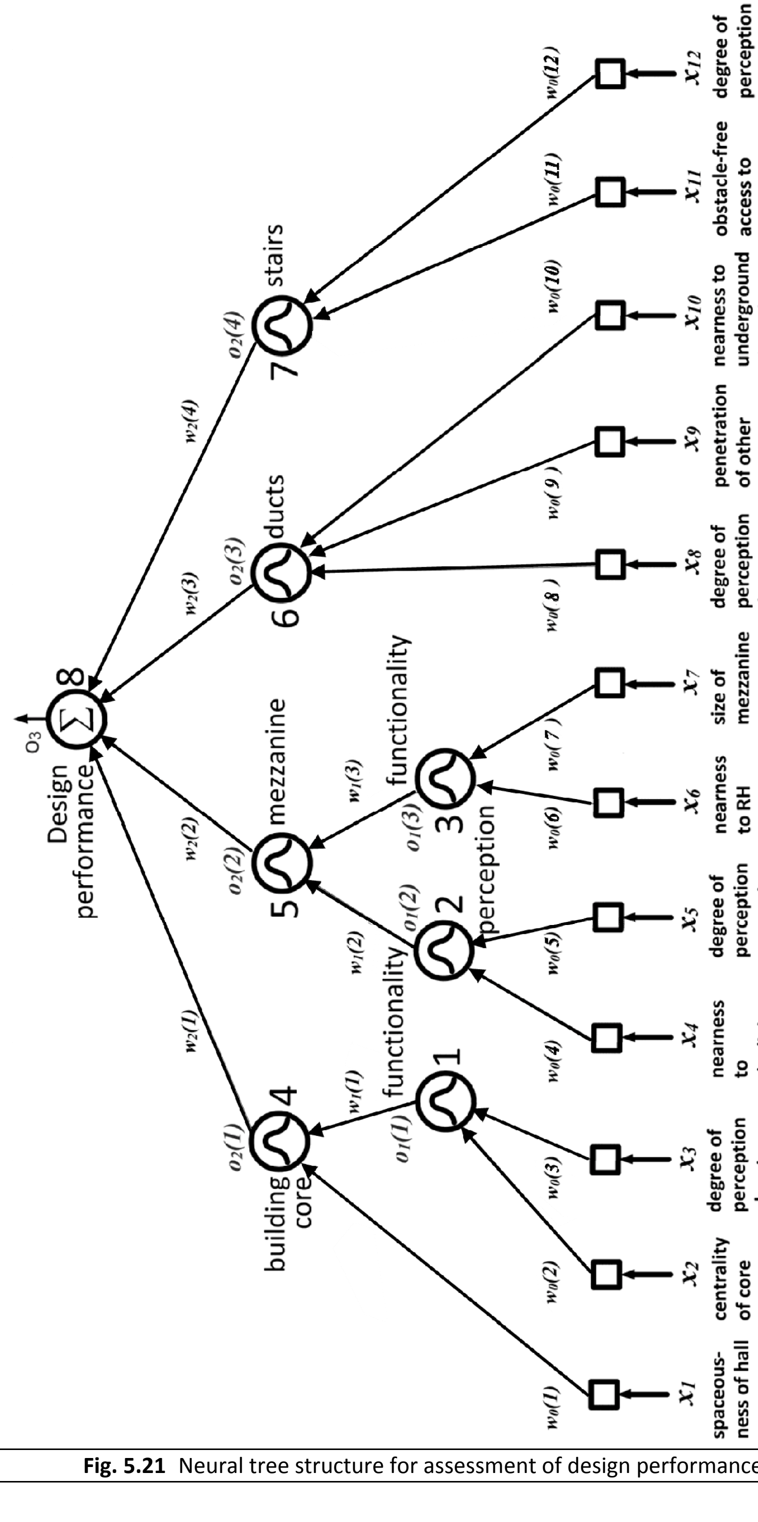


Fig. 5.21 Neural tree structure for assessment of design performance

Table 5.9 Determinants of the design performance

building core	mezzanine	ducts	stairs
spaciousness of the entrance hall	nearness to daylight	degree of perception of the ducts	obstacle-free access to elevators
centrality of the building core	degree of perception of the mezzanine	No penetration of other objects	Degree of perception of the stairs
degree of perception of the elevators	nearness to adjacent space	nearness to underground outlet	
	size of the mezzanine		

Thereafter the outcomes of the neural tree are investigated in the framework of multi-objective optimization. In the context of the design application the design performance is determined by four sub-domains. These domains are the performance scores of the design objects. The scores are given at the node outputs on the penultimate level of the tree. At one level further below is the terminal level. This is except with respect to the performance of the building core and the mezzanine, where the performance has two additional sub-aspects. For the building core the sub-aspect is functionality; and for the mezzanine the sub-aspects are its perception and functionality.

The determinants of the design performance on the terminal level are given in table 5.9. These are the elemental design requirements. The specification of these requirements is provided item-wise below. The corresponding fuzzy membership functions are shown in figures 5.22-5.31. It is noted that load-bearing aspects are excluded in the demonstrative exercise for simplicity. Due to the transparent structure of the neural tree concept, future integration of such additional requirements is straightforward and an interesting relevance.

5.2.3. Modelling the requirements at the terminal nodes

The requirements considered in the application are as follows.

Spaciousness of the entrance hall refers to the volume of the entrance hall that spans from the floor to the ceiling of the building and from the observer to the building core, while excluding the portion that is covered by the mezzanine.

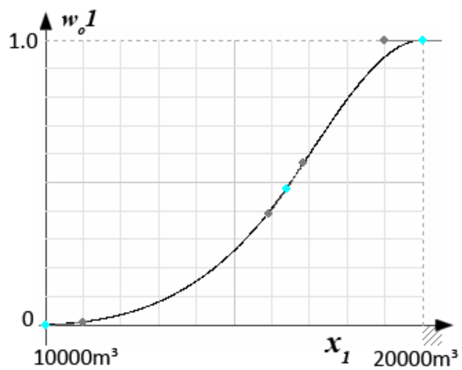


Fig. 5.22 MF expressing the requirement on the spaciousness of the entrance hall

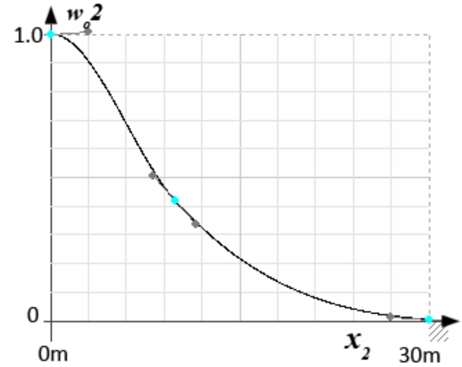


Fig. 5.23 MF expressing the requirement on the centrality of the building core

The requirement is expressed by the fuzzy membership function shown in figure 5.22. From the figure it is seen that more spaciousness is considered better for the performance and that the requirement is fully satisfied beyond a volume of $20000m^3$, while it is considered not satisfied at all below $10000m^3$.

Centrality of the building core refers to the distance of the building core from the center of the space. The distance is measured from the center point of the building core to the center point of the hall. This is expressed by the membership function shown in figure 5.23. From the figure it is clear that more distance is negative for the performance. This requirement is motivated by the consideration that the central location is most equally accessible from any place, which is desirable for the accessibility of the parking garage and the offices in the tower. This may strengthen the functionality of the entrance hall as a *hub*, distributing people flow. From the figure it is seen that the requirement is fully satisfied when the building core is exactly at the center location, i.e. the distance is zero. The requirement is considered not satisfied at all at a distance beyond 30m.

Degree of perception of the elevators refers to the visual perception of the elevators that are located in the centre of the building core from the location of the observer.
The motivation of the requirement is to ensure that the elevators are easily noticed from the entrance of the hall, so that visitors of the offices have a rather effortless experience reaching their destination. This is expressed by the fuzzy membership function shown in

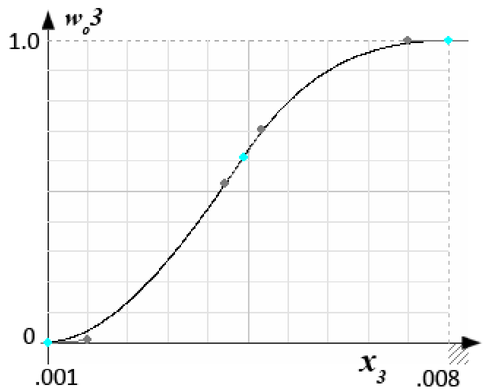


Fig. 5.24 MF expressing the requirement on the degree of perception of the elevators

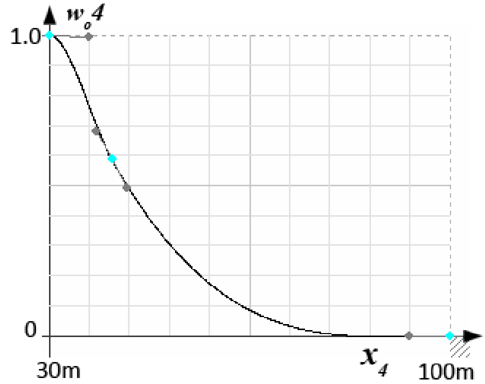


Fig. 5.25 MF expressing the requirement on the nearness to daylight

figure 5.24. From the figure it is clear that more perception is positive for the performance, and the requirement is considered fully satisfied beyond a perception of 8.0 10^{-3}. The meaning of this number is illustrated considering perception of a square patch of $1m^2$, where its center is located directly at eye height and 20m in front of an observer. The scope of the observer's vision is taken as 180°. The degree of perception of this patch is obtained to be 2.6 10^{-3} using the probabilistic perception theory in chapter 2.

Nearness to daylight refers to the distance between the mezzanine and the glass façade of the space. The distance is measured from the centre points of the two objects.
The distance is desirably small as to benefit from daylight on the mezzanine. This is expressed by the membership function in figure 5.25. The requirement is considered fully satisfied for a distance less than 30m, whereas it is not satisfied at all for distances beyond 100m.

Degree of perception of the mezzanine refers to the degree of visual perception of the mezzanine viewed from the location of the observer. The degree is desirably small to avoid that the mezzanine dominates the space. This is expressed by the membership function in figure 5.26. The requirement is considered fully satisfied for a degree of perception below 0.05, whereas it is not satisfied at all for a degree beyond 0.3.

Nearness to adjacent space refers to the distance between the mezzanine and an

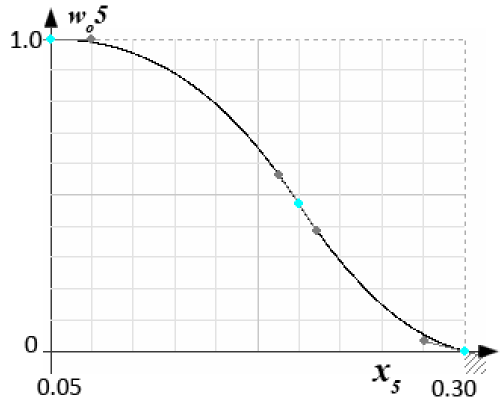


Fig. 5.26 MF expressing the requirement on the degree of perception of the mezzanine

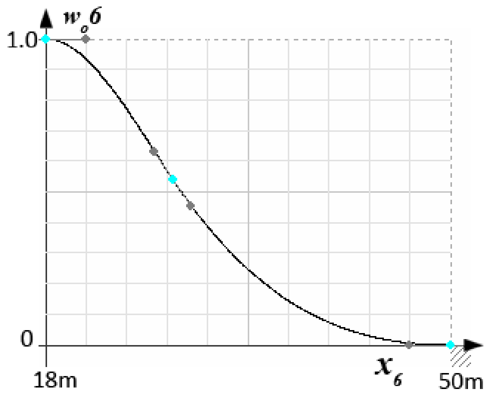


Fig. 5.27 MF expressing the requirement on the nearness to the adjacent space

adjacent space *named "Rotterdam Hall."* The distance is measured from the centre point of the mezzanine to the entrance of the space. The distance is desirably small, so that the Rotterdam hall may be accessed via the mezzanine providing additional accessibility to this space. This is expressed by the membership function in figure 5.27. The requirement is considered fully satisfied for a distance less than 18m, while it is considered not satisfied at all for a distance beyond 50m.

Size of the mezzanine is a self evident description. The size is required to be big enough for its designated use as a restaurant, yet not too big, as this would not be needed currently, while it would be expensive. The latter restriction is less stringent compared for the demand for sufficient size, because the extra space provides extra flexibility. This is expressed by the fuzzy membership function in figure 5.28. From the figure it is seen that the requirement is considered fully satisfied for a size between 220m^2 and 250 m^2. The mezzanine is considered too small at a size below 100m^2. The requirement is considered to be fulfilled by 50% for a size beyond 500 m^2.

Degree of perception of the ducts refers to the visual perception of the ducts from the viewpoint of the observer. There are two ducts. Their position is constrained to be

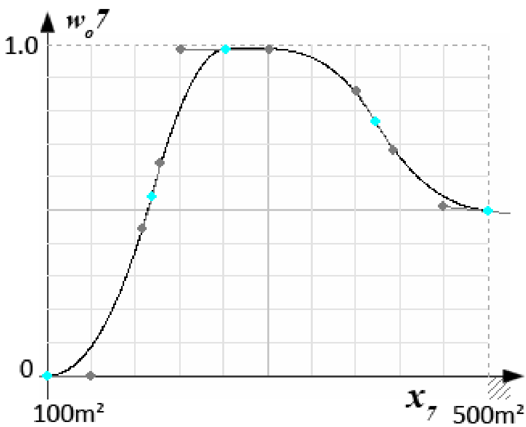


Fig. 5.28 MF expressing the requirement on the size of the mezzanine

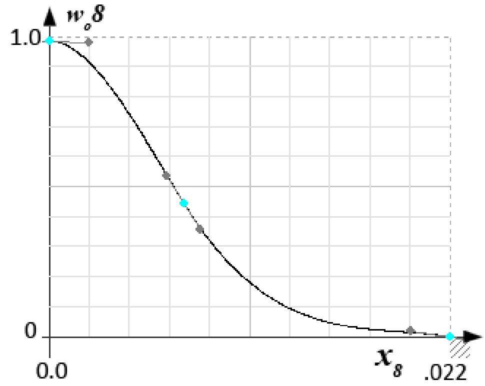


Fig. 5.29 MF expressing the requirement on the degree of perception of the ducts

symmetrical to the axis passing from the elevators in the forward direction of the observer. It is desirable that the perception of the ducts is low. This is expressed by the fuzzy membership function shown in figure 5.29. From the function it is seen that the requirement is considered fully satisfied for a degree of perception nearly zero. It is considered not satisfied at all for a perception greater than .02.

No penetration of other objects refers to the event that none of the shafts is penetrating the mezzanine or the stairs, or that they are forming obstacles. The intention behind this is to avoid additional costs in the construction, reduced functionality and compromised aesthetic appeal. If a shaft penetrates the stairs, satisfaction of this requirement becomes zero. If it penetrates the mezzanine satisfaction is considered to be .3. If the column is obstructing the direct passage from the observer location to the elevators, satisfaction of the requirement becomes .2. Thus, penetration of the mezzanine is considered less severe than forming an obstacle in the movement trajectory. It should be stressed that the degree of perception of the ducts may be manipulated by both position and colour, introducing greater degrees of freedom in the search.

Nearness to the underground outlet refers to the distance between the ducts and an outlet location for an existing underground HVAC system. The distance is desirably small to

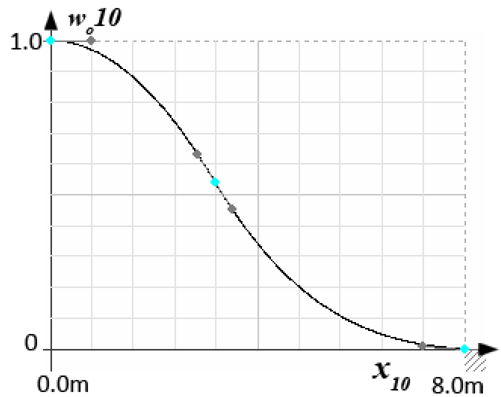


Fig. 5.30 MF expressing the requirement on the nearness to the outlet

avoid losing space in the parking garage due to horizontal movement of the underground ducts. This is expressed by the fuzzy membership function shown in figure 5.30. From the figure it is seen that the requirement is fully satisfied for a distance of zero and not satisfied at all for a distance greater than 8m.

Obstacle-free access to elevators refers to the situation that the stairs do not impede the access to the elevators from the entrance. To identify the obstruction conveniently, two virtual corridors are placed, so that both connect the observer location and the elevators with their central axis. The corridors' have their axis on the same line, while they have different widths. The wider corridor of the two is 8m wide and 2.5m high, and the narrow one is 2.20m wide and 2.10m high. Testing the collision of the stairs with the corridor objects, two different degrees of path obstruction can be distinguished. The narrow corridor approximately represents the minimal space necessary to comfortably walk to the elevator in a straight line. Therefore penetration of the narrow corridor by the stairs is considered to reduce the satisfaction of this requirement to zero. Penetration of the wider corridor may be problematic in case the space is crowded. Therefore it is considered to reduce the satisfaction to 0.6. The penetration is identified by collision tests.

Degree of perception of the stairs refers to how strongly the stairs will be noticeable from the entrance. The degree of perception is desirably not too big in the present case, because the mezzanine is considered to be a secondary space with a private character, and therefore its access should not be made prominent giving the impression the mezzanine were a public space. From the other side the stairs should be sufficiently noticeable from the entrance, not to make it too difficult to find them when needed. This is expressed by the fuzzy membership function shown in figure 5.31. From the figure it is seen that the ideal degree of perception of the stairs is considered to be around .02, while the satisfaction of the requirement vanishes for perception beyond 0.10.

It should be noted that the shape of the functions can be interpreted in relation to the *tolerance* for the full satisfaction of the requirement: A wider shape expresses a greater tolerance than a narrow shape. In a narrow case slight deviation from the ideal situation is yielding drastic reduction in requirement satisfaction, while in the wide case slight deviation has little effect on the requirement satisfaction. As it is seen from the figures, the physical units of the dimensions considered in the requirements differ among requirements, and the degree of perception is unit-less.

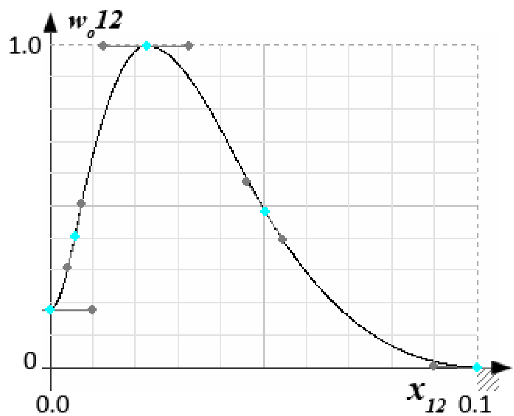


Fig. 5.31 MF expressing the requirement on the degree of perception of the stairs

Table 5.10 Weights of the neural tree for evaluation of the design performance

weight nr.	1	2	3	4	5	6	7	8	9	10	11	12
level 1	.60	.70	.55									
level 0	.80	.40	.70	.50	.70	.40	.80	.70	.40	.60	.40	.80

Therefore the fuzzy membership function defining the satisfaction of the requirements between zero und unity can be seen as *standardization* bringing the different quantities on a common ground for further computation.

The determinants shown in table 5.10 form a multidimensional search space, which is complex with respect to its dimensionality. In this space, Pareto optimality is most desirable for multi criteria based search. For the tree structure established, the connection weights at each level are assessed by an architect and given in table 5.10. These weights indicate the relative importance of a sub-aspect compared to other sub-aspects. The structure can be considered as constituting the domain knowledge, where the connecting weights between the nodes are determined by expert judgment. This means there is no formula or computational mechanism used to establish the weights. The weights and the neural tree structure are an expression of an expert knowledge on what are the significant aspects that make the interior a suitable for its intended purpose, in particular defining which aspects determine the performance of the design objects, and their relative significance. It is to be noted that level 2 is not specified in table 5.10, since in the present case the relative importance among the design objects is problematic to specify for a designer before having an impression on the nature of the conflicts among design objects. This means it is left up to the computational cognition to hint designers what relative importance associated with the objects yields maximal design performance in the present task.

As far as non-terminal nodes are concerned the widths of the Gaussians are still to be determined. These are obtained by means of the consistency condition, which serves as boundary condition for the neural tree model.

5.2.4. Imposing the consistency condition

The imposition of the consistency or boundary conditions can be carried out by adaptive or genetic learning.

Table 5.11 Adaptive learning results from the imposition of the consistency condition

input at the leaf nodes	approximation at the output of the inner nodes on level 2				error			
	core	mezza-nine	ducts	stairs	core	mezza-nine	ducts	stairs
0.100	0.197	0.184	0.109	0.106	-0.097	-0.084	-0.009	-0.006
0.200	0.243	0.222	0.174	0.170	-0.043	-0.022	0.026	0.030
0.300	0.295	0.271	0.262	0.257	0.005	0.029	0.038	0.043
0.400	0.360	0.337	0.374	0.369	0.040	0.063	0.026	0.031
0.500	0.450	0.437	0.505	0.500	0.050	0.063	-0.005	0.000
0.600	0.582	0.586	0.646	0.642	0.018	0.014	-0.046	-0.042
0.700	0.752	0.776	0.782	0.779	-0.052	-0.076	-0.082	-0.079
0.800	0.906	0.933	0.897	0.895	-0.106	-0.133	-0.097	-0.095
0.900	0.984	0.995	0.973	0.973	-0.084	-0.095	-0.073	-0.073

Table 5.12 Mean squared errors for the outputs from consistency imposition

object	core	mezzanine	ducts	stairs
error	4.08E-03	5.45E-03	2.94E-03	2.90E-03

As result of the learning process, the width of each individual Gaussian at each non-terminal node is established. In this way, the cascade feed-forward fuzzy logic operations are clearly defined exhibiting features of transparency in the model. In the present implementation adaptive learning is used for high accuracy. The computation at the root node is a defuzzification of the fuzzy information processed at the inner nodes. It is in the form of a weighted summation. Therefore the consistency condition may be imposed separately on the four branches of the tree. This means the consistency for each of the four branches is established separately.The approximation error for this data set is relatively higher for the lower and higher input/output pairs. This is seen from table 5.11, where the results from the learning are shown. The mean square errors are shown in table 5.12 and the resulting widths of the Gaussians at the terminal nodes are shown in table 5.13.

Having established the fuzzy neural tree, the task is to maximize the output at the root node by identifying optimal parameters of the design objects. This is accomplished by genetic search due to the complexity of the search domain. The output at the root node expresses the design performance by a scalar number. This value can be used as the representation of the fitness of the respective chromosome. In this way the genetic algorithm (GA) would search for solutions with maximal performance, however it would be treated as a single-objective problem.

However, in design we are unable to specify in advance the relative importance among the major design aspects. This is because we are unaware of the precise implications of such commitment. In the present case it is problematic to pin-point how important the design objects should be considered for the overall design performance. How much importance shall be assigned to the stairs for example compared to the building core? The difficulty stems from the fact that the objects have multiple, conflicting requirements attached, and the implications of their simultaneous satisfaction is unclear. Therefore the problem is treated as a multi-objective problem. This way it is not necessary to commit prematurely regarding the relative importance among the criteria.

5.2.5. Generating Pareto-optimal designs

In the present application four criteria are to be maximized simultaneously. These are maximization of the respective performance of the four objects forming the design. Therefore, the Pareto-optimal frontier is in a four dimensional space. This space is formed by the outputs on the penultimate level of the neural tree.

Table 5.13 Resulting widths of the Gaussians at the non-terminal nodes

Node nr.	1	2	3
σ	$2.0\ 10^{-1}$	$5.0\ 10^{-1}$	$2.3\ 10^{-1}$

Node nr.	4	5	6	7
σ	$5.2\ 10^{-1}$	$4.0\ 10^{-1}$	$4.3\ 10^{-1}$	$3.8\ 10^{-1}$

Table 5.14 hierarchy determining positional dependence among the objects

building core	
mezzanine	ducts
stairs	

It is emphasized that design objects usually fulfil certain basic conditions. For example stairs are usually connecting two floors. Such basic relations should be invariant for any solution we consider. Therefore it is appropriate to restrict the evolutionary search, so that it only considers solutions that are feasible in this sense.In the present application such basic relations concern geometric dependence among objects. For example, the range of possible positions of the stairs depends on the position of the mezzanine, as the stairs naturally connect the mezzanine with the ground floor level. This basic relation should always prevail for any solution, i.e. it is invariant.In the present application this invariance is implemented by dynamically changing the manoeuvring area of depending objects, which is done by adjusting the search boundary of the genetic algorithm. For example, as the mezzanine is displaced, the range for possible positions of the stairs is also displaced along with the mezzanine. The dependence among objects forms a hierarchy, where the location of objects on the lower level of the hierarchy depends on the location of an object on the higher level. The hierarchy is shown in table 5.14. From the table it is seen that the stairs are depending on the mezzanine, while the mezzanine is depending on the position of the building core. The ducts are also depending on the position of the core. In the present application the evolutionary algorithm controls the position of lower level objects *relative* to the position of the higher level object, and not relative to the origin of the scene.

To identify the Pareto frontier a multi-objective GA is employed using Pareto ranking as fitness criterion. This means the fitness of a solution R_{fit} in a population depends on the amount n of other chromosomes that are dominating it in the objective space. The fitness is obtained by

$$\boldsymbol{R}_{fit} = \frac{1}{N(\theta)+n} \cdot \tag{5.11}$$

Thereafter the fitness is converted to a probability for reproduction applying the well-known roulette wheel selection principle [13].

Preserving diversity among solutions on the Pareto front is desirable. This is so that we are able to select among the solution alternatives with great awareness. To alleviate the common problem of aggregating solutions on the Pareto front, a *relaxed* dominance concept is employed as described in Chapter 4. This means the scope within which the degree of non-dominance of a solution is measured is increased compared to the conventional case, where it is orthogonal. In this way solutions that are conventionally not considered to dominate a particular solution under concern are now being considered to be dominant. This entails that potential solutions are not prematurely excluded during the search, so that the front is developed carefully. The widening of the scope of tolerance is implemented by means of coordinate transform of the solutions.

For this implementation we design the F matrix in Eq. (4.7) in Chapter 4 as

$$\begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \end{bmatrix} = \begin{bmatrix} 1 & -.176 & -.176 & -.176 \\ -.176 & 1 & -.176 & -.176 \\ -.176 & -.176 & 1 & -.176 \\ -.176 & -.176 & -.176 & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} \tag{5.12}$$

This implies that 10^o are employed between the respective orthogonal and non-orthogonal system axes, so that, the cosine direction matrix given by (4.14) in Chapter 4 becomes

$$D = \begin{bmatrix} .956 & -.168 & -.168 & -.168 \\ -.168 & .956 & -.168 & -.168 \\ -.168 & -.168 & .956 & -.168 \\ -.168 & -.168 & -.168 & .956 \end{bmatrix} = Q^T \qquad (5.13)$$

The corresponding weighted objectives F_1 and F_2 are given by

$$\begin{bmatrix} f_1^P \\ f_2^P \\ f_3^P \\ f_4^P \end{bmatrix} = \begin{bmatrix} .956 & -.168 & -.168 & -.168 \\ -.168 & .956 & -.168 & -.168 \\ -.168 & -.168 & .956 & -.168 \\ -.168 & -.168 & -.168 & .956 \end{bmatrix} \begin{bmatrix} F_1^P \\ F_2^P \\ F_3^P \\ F_4^P \end{bmatrix} \qquad (5.14)$$

and the inverse of Eq. (5.13) becomes

$$\begin{bmatrix} F_1^P \\ F_2^P \\ F_3^P \\ F_4^P \end{bmatrix} = \begin{bmatrix} 1.220 & .331 & .331 & .331 \\ .331 & 1.220 & .331 & .331 \\ .331 & .331 & 1.220 & .331 \\ .331 & .331 & .331 & 1.220 \end{bmatrix} \begin{bmatrix} f_1^P \\ f_2^P \\ f_3^P \\ f_4^P \end{bmatrix} \qquad (5.15)$$

Eq. (5.14) is the coordinate transform to verify dominance among solutions using the relaxed approach.

Figure 5.32 shows a design at the beginning of the genetic search process, which is a random configuration. It has a low design performance, i.e. the output value at the root node O_3 of the tree is low.

There are several reasons for this. A major issue is that the entrance hall is not as spacious as required, since the building core is positioned too close to the entrance. A second issue concerns the ducts, which are undesirably close to the avatar, so that they visually dominate the space too much. The same counts for the stairs. Their perception is higher than required. Next to this the stairs and duct2 are intersecting each other causing additional construction costs. The result of the search process is shown in figure 5.33, where

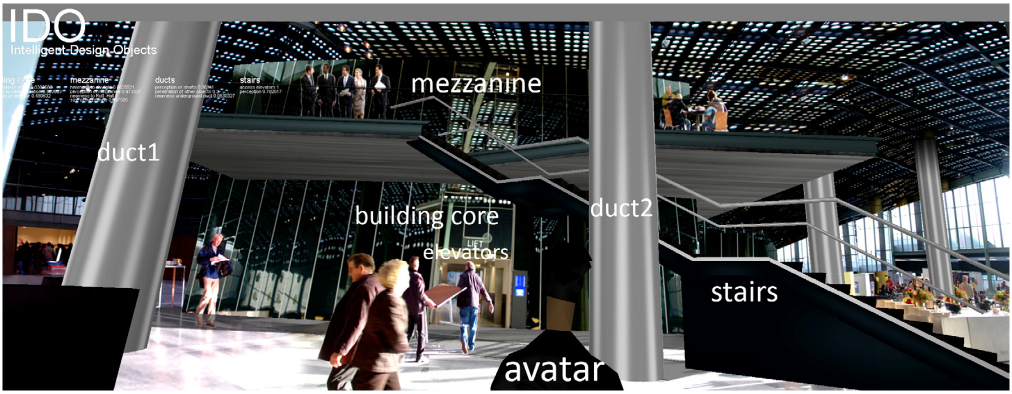


Fig. 5.32 Design at the beginning of the Genetic search process with a performance of 0.45

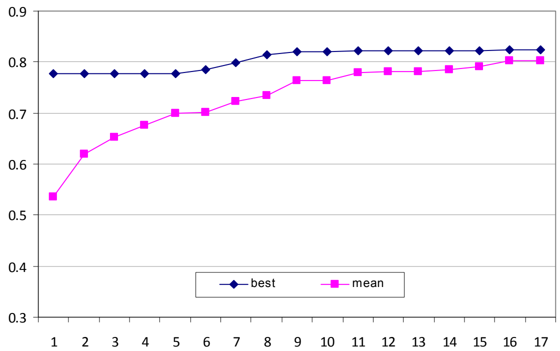


Fig. 5.33 Genetic search results with respect to the design performance

the best fitness that occurred during the search is plotted together with the average fitness of the chromosomes for each generation. It is noted that the fitness for the best solutions shown in the figure is the output at the root node of the neural tree, and it is used to indicate the convergence of the search, while the output of the root node has no relevance in the multi-objective search.

From figure 5.34 it is seen that the difference among the solutions of the final population in terms of design performance is smaller than compared with the initial population. Most solutions in the final population have a design performance close to .80. For the initial population clearly the performance varies much greater, while the average performance is around .55.

5.2.6. Resulting Pareto-optimal designs

After 17 generations the population of the GA arrived at the Pareto frontier. From figure 5.34 it is seen that after the search the solutions on the Pareto front have similar maximal performance scores, whereas in the beginning of the search the scores of the solutions differ greatly, and they are significantly lower as well compared to the Pareto solutions.
The resulting designs are equally valid solutions in Pareto-sense, while each solution has a different trade-off with respect to the design criteria. The results from the search provide means to comprehend the trade-offs that characterize the Pareto optimal alternatives. This way a decision maker obtains new information in order to validate the criteria. Figure 5.35 shows the Pareto optimal designs obtained after the multi-objective search using *greedy* Pareto ranking.. Figure 3.36 shows the Pareto solutions using *relaxed* Pareto ranking.

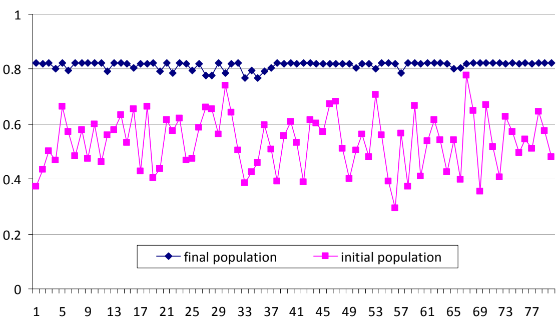


Fig. 5.34 Performance of the solutions in the 17th generation

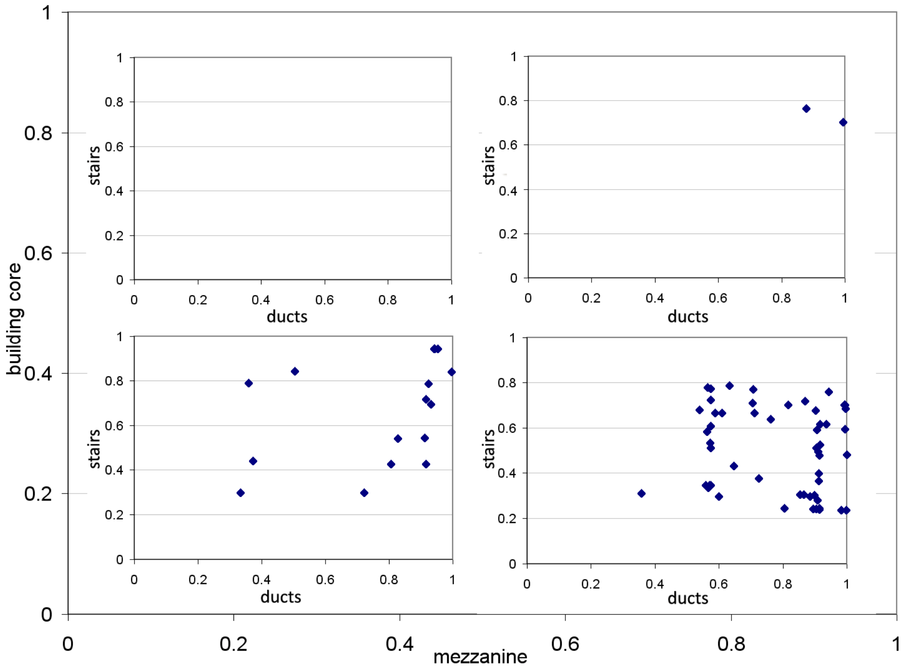


Fig. 5.35 Pareto optimal designs with respect to the four objective dimensions using greedy Pareto ranking

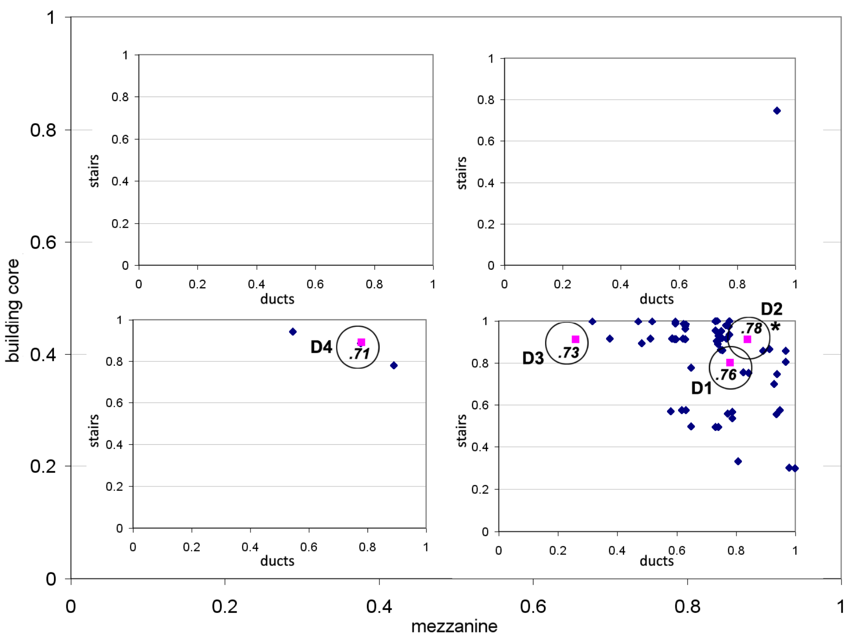


Fig. 5.36 Pareto optimal designs with respect to the four objective dimensions using relaxed Pareto ranking

It is noted that the solutions in figures 5.35 and 5.36 are plotted in the objective space, which has four dimensions, one for every design object. For the multi-dimensional representation the solutions are first divided into four groups, depending on which quadrant of the mezzanine/building core space they belong to respectively. Then a solution is shown within its respective quadrant based on its stairs-ducts performance score. This form of representation is used, since it has advantages compared to the well known scatter plot [16] and value plot diagram [17]. Scatter plots show the solutions with respect to two objective dimensions at a time. The two-dimensional representation does not give a clue about the location of a point with respect to the other dimensions at the same time. In the representation used here, due to the subdivision of two objective dimensions into sub-domains, when a two-dimensional position of a point is inspected, at the same time the context, that is, the location of this very point with respect to the other two objective dimensions is comprehended at some resolution at the same time. Value plot diagrams are useful only to identify aggregation on the Pareto front. Since value paths show every objective dimension separately the shape of the Pareto front, i.e. the relative locations of the points in a multidimensional space is difficult to comprehend, as the space should be comprehended as a multidimensional unit.

It is noted that the scores shown in figure 5.34 are the maximal performance obtained using Eq. (5.16).

It is interesting to compare the relaxed dominance concept used in the Pareto ranking with its greedy counterpart. For this purpose the result from a greedy search in figure 5.35 are obtained after the same number of generations as in figure 3.36 showing the results from a relaxed search. The positive effect of relaxation is observed comparing the Pareto fronts in figure 3.35 and 3.36. In the greedy case the front did not establish very clearly. This is seen from figure 5.35, where some solutions that are inferior in ducts/stairs objective space remain in the population, while these are not present in the relaxed case. We can say that the solutions in the relaxed case are more 'motivated' to move towards the front compared to the greedy case. This is explained considering that in the four dimensional objective space the size of the space is large that merely few solutions are dominated in a strict, i.e. greedy, sense. This results in low pressure towards the Pareto front. In the relaxed

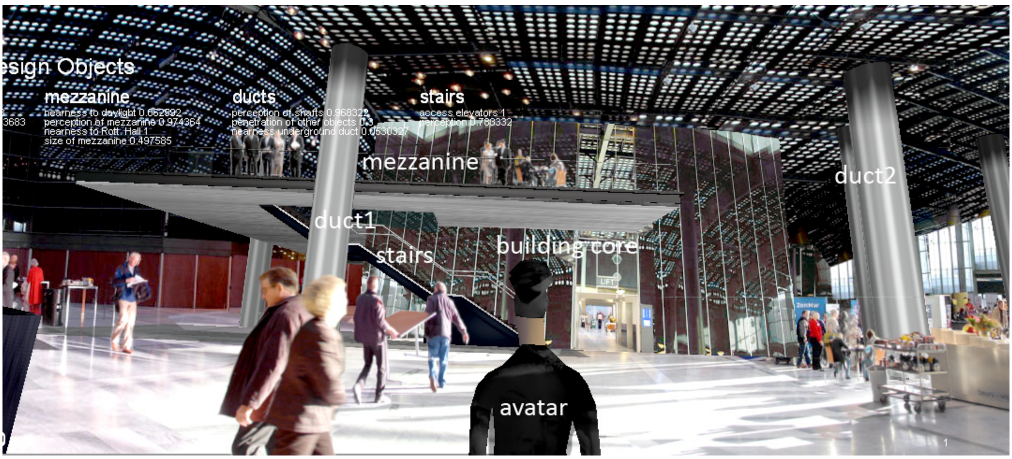


Fig. 5.37 Design *D1*

case some inferior solutions are counted as being dominant, due to the wider scope of tolerance. In this way the selection pressure is finely adjusted. The potential of a greater portion of the population is contributing to the front formation, since they are not prematurely excluded during the search. In figure 5.36 four designs, D1-D4, are highlighted, and their maximal performance scores obtained using Eq. (5.16) are indicated as well. The corresponding scenes are shown in figures 5.37 – 5.40. The performance scores of the design objects that characterize these four designs are shown in figure 5.41.

From the scenes it is seen that the Pareto solutions satisfy the requirements put forward in a general sense, so that the Pareto optimal solutions have some common features. In all solutions the entrance space is large, while the elevator access still has a relatively high level of perception. This is because the elevator entrance is located directly in front of the avatar. The ducts are generally located at a large distance from the avatar reducing their perception. At the same time they do not penetrate the mezzanine or the stairs, or obstruct the passage to the elevators. Considering the differences among the solutions we note that each solution has a specific characteristic regarding the performance of the design objects. That is, in one solution a certain object performs better at the cost of the performance of other ones. This difference among the four designs is highlighted in figure 5.41.

Comparing D1 with D2 we note that in D1 the building core and the mezzanine perform better than in D2. This is because in D2 the building core is located nearer to the avatar. This reduces the spaciousness of the hall undesirably, this conflicting with a requirement. At the same time the mezzanine is located nearer to the avatar in D2 compared to D1. Therefore in D2 the perception of the mezzanine becomes undesirably high. This reduces the performance of the mezzanine in D2 with respect to D1. The superior performance of the building core and the mezzanine in D1 is at the cost of lower performance of the ducts and stairs in D1 compared to D2. This is because the nearness of the building core to the avatar implies that the stairs and ducts are near as well. Therefore their positioning is favourable in some respect. Namely, as the position of the stairs is closer to the avatar in D2 compared to D1, their perception is higher in D2. This is desirable due to the corresponding requirement. Concerning the ducts, the distance from the underground outlet is lower in D2 compared to D1. Therefore D2 outperforms D1 in this respect.

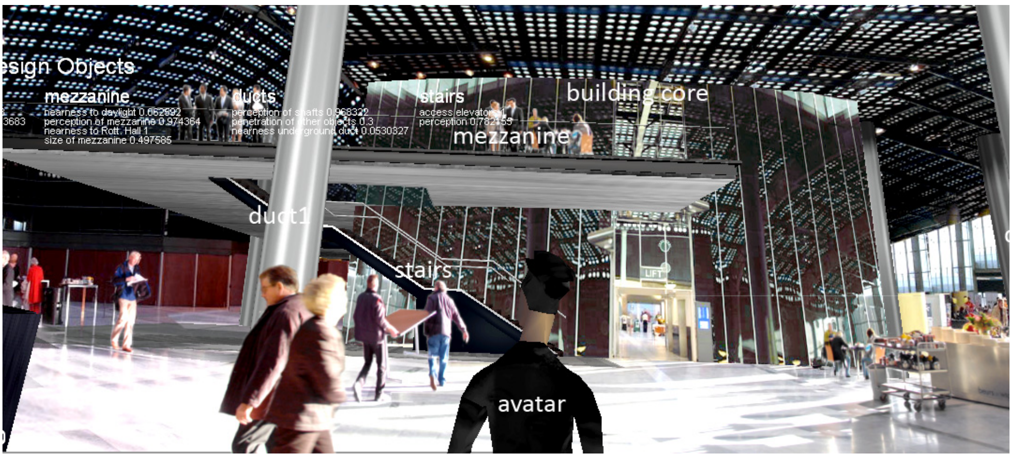


Fig. 5.38 Design *D2*

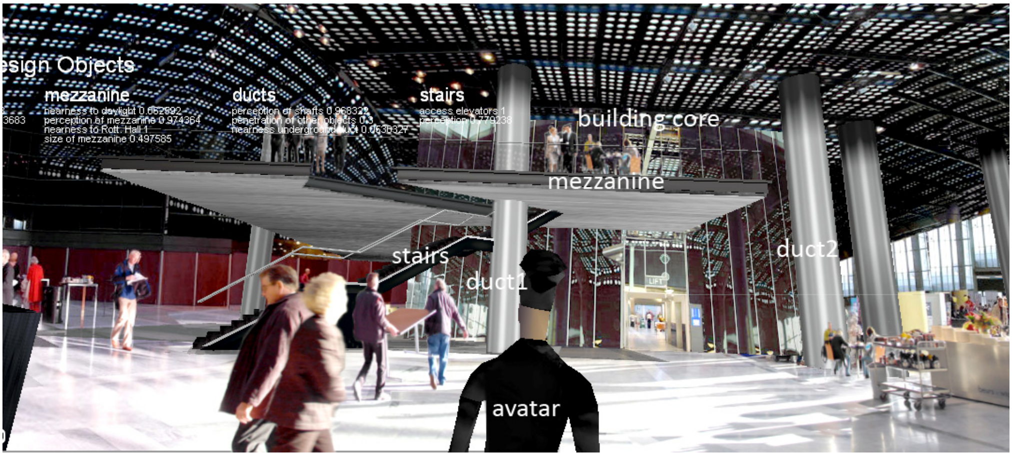


Fig. 5.39 Design *D3*

Comparing D1 and D3 we note that these designs differ with respect to the performance of the stairs. This was also the case comparing D1 and D2. Since the stairs in D3 have a desirably higher degree of perception, D3 is superior in this respect. It should be noted that D3 is particular in the sense that it has a low performance of the ducts. This is because a duct penetrates the stairs, which is undesirable according to the requirements. Also the ducts in D3 are located distant from the underground outlet, so that their performance is further reduced.

Comparing D3 and D4 we note that in D4 the ducts are not penetrating the stairs, so that their performance is higher in this case compared to D3. However, at the same time the mezzanine in D4 has a lower performance compared to D3. This is because in D4 the mezzanine is located very far from the adjacent space, which is undesirable according to the corresponding requirement. Initially it is problematic to specify the relative importance among the design objects precisely. After the evolutionary search, having a large repertoire of Pareto optimal solutions we are in a position to specify this importance. First we note that a among the Pareto solutions a certain solution yields the maximal

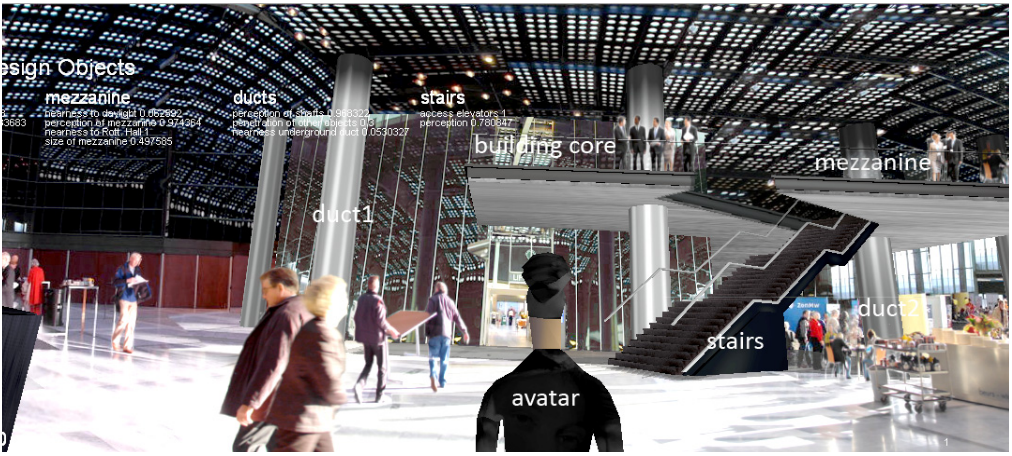


Fig. 5.40 Design *D4*

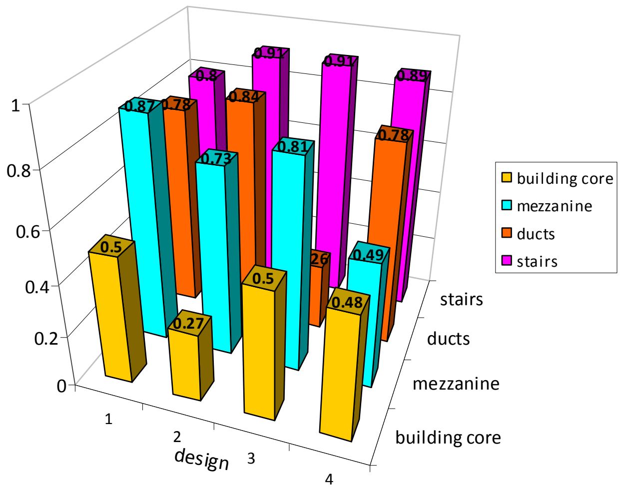


Fig. 5.41 Comparison among designs D1-D4

performance score when we exercise any possible combination for the weights on the penultimate level.

This solution is obtained by computing the maximal performance score for every solution as given by

$$p = \frac{O_1(1)^2 + O_2(1)^2 + O_3(1)^2 + O_4(1)^2}{O_1(1) + O_2(1) + O_3(1) + O_4(1)} . \qquad (5.16)$$

and selecting the solution with the highest p value among them. Eq. (5.16) implies a preference vector ***w**** for the weights at the penultimate tree level. Modifying the preference vector we are able to explore the solutions on the Pareto front to identify a solution suiting our preferences most closely. This means we may select any of the designs by adjusting the weights on the penultimate level of the neural tree. Based on this higher-level criterion the intelligent objects present a certain design located on the Pareto front, which is both non-dominated and has the highest output value at the root-node of the neural tree fulfilling our higher-order demand.

It is noted that in this process we can exercise our preferences with increased awareness compared to the case we design using conventional means. This is because any solution on the Pareto front already fulfils the primary objectives we put forward, so that we can focus entirely on the second-order preferences. Using conventional means of design we are not free to do so. This is because we have to also take care of the complex relations among the design objects next to exercising higher-order preferences.

In the present design task the design with the highest score p among the Pareto solutions highlighted is solution D2 having a score of .78, followed by D1 scoring .76; D3 has .73 and the lowest score belongs to D4, namely .71. This shows that in the present design task solutions in the lower right quadrant of the mezzanine building core space are preferred over solutions in the lower left quadrant. This means that in the present case it is preferable to have the mezzanine on the left hand side of the avatar, so that the mezzanine performance is high, and at the same time to ensure that the ducts do not penetrate the stairs. It is emphasized that this insight is newly obtained from the inspection of the Pareto

Table 5.15 Comparison of an initial and Pareto-optimal design D2

Node	Output	Initial design	Pareto-optimal design D2
Design performance	O_4	.40	.78
Perf. Building core	$O_3(1)$	.23	.27
Perf. Mezzanine	$O_3(2)$	.32	.73
Performance Ducts	$O_3(3)$	.43	.84
Performance Stairs	$O_3(3)$	.83	.91
Functionality of building core	$O_2(1)$	.02	.00
Perception of mezzanine	$O_2(2)$	.28	.55
Functionality of mezzanine	$O_2(3)$	.39	.99
Spaciousness entrance hall	w_1	.17	.25
Centrality core	w_2	.83	.83
Elevator perception	w_3	.20	.23
Nearness mezzanine from daylight	w_4	.26	.22
Perception mezzanine	w_5	.00	.46
Nearness mezzanine from "Rotterdam Hall"	w_6	.86	.98
Spaciousness mezzanine	w_7	.61	.96
Perception ducts	w_8	.95	.64
Penetration of objects by ducts	w_9	.30	.99
Nearness ducts from underground hub	w_{10}	.81	.99
Unobstructed access elevators from entrance	w_{11}	.60	.60
Perception of stairs	w_{12}	.78	.95

front based on the results from computational cognition, and did not exist prior to the design task or could not have been obtained otherwise.

The node outputs for an exemplary Pareto solution are shown in table 5.15 together with the outputs for an initial design for comparison. From the table it is noted that compared to the random design the Pareto optimal one performs superior regarding most requirements.

5.2.7. A variation of the design task

In order to illustrate the consistency of the intelligent object's behaviour satisfying the requirements, the approach is employed to handle a modified version of the previous design task. The modification concerns the required degree of perception of the stairs. Previously it was demanded that the stairs should not be very much noticeable from the

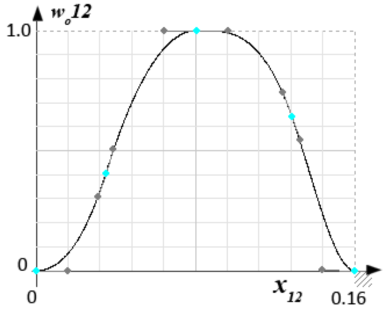


Fig. 5.42 MF expressing the requirement on the degree of perception of the stairs

entrance. In the present case it is demanded that the stairs are more intensely noticeable. This requirement corresponds to a scenario where the intended use of the mezzanine is retail. In a retail environment customers should become strongly aware of the fact that there is a second floor with shopping articles. This is expressed in the fuzzy membership function shown in figure 5.42. Comparing the new requirement with the previous one we note that the newly required degree of
perception is much greater than before. The satisfaction of the requirement in the previous case vanishes after a degree of perception of 0.1, whereas in the present case the maximum satisfaction is around that very amount of perception. A second modification concerns the significance of the perception of the stairs. In the present design task it is considered more relevant than in the previous case. Previously the significance of the perception of the stairs was .8. In the present case the significance is increased to .9. One notes that the weight values connecting to an inner node need not necessarily sum up to one. This is taken care of by properly adjusting the width of the Gaussian at the node.

It is noted that when the neural tree is merely changed at the weights between root node and one level below, there is no need to re-establish the consistency condition. So the change is immediately effective. However in this variation of the design task the weights connecting a terminal to an inner node are changed. Therefore it is necessary to re-establish the consistency condition of the tree. In particular the width at node 7 is to be re-established.

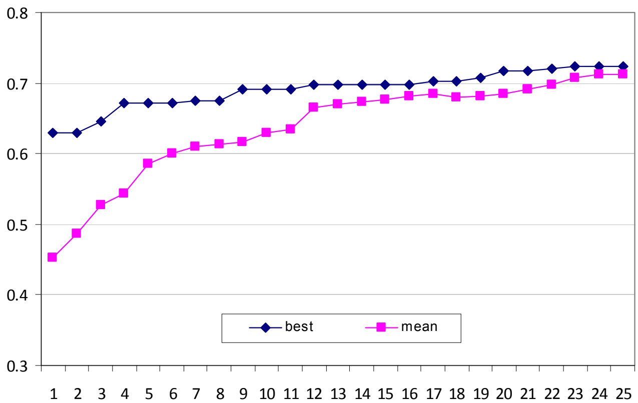


Fig. 5.43 Genetic search results with respect to the design performance

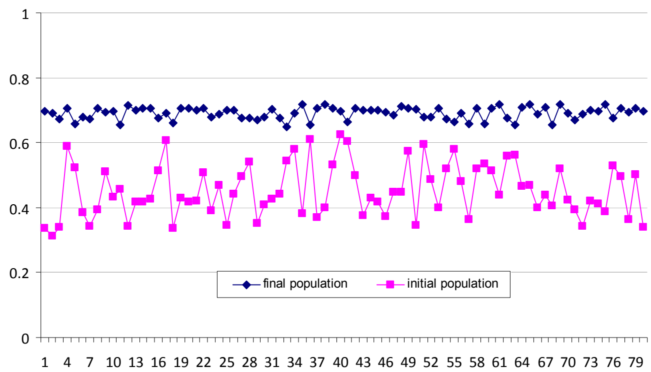


Fig. 5.44 Performance of the solutions in the beginning of the search and after 20 generations

The result of the search process is shown in figure 5.43, where the best fitness occurring during the search is plotted together with the average fitness of the chromosomes for every generation. From figure 5.44 it is seen that after 20 generations the Pareto optimal solutions have a similar performance score, while the scores strongly differ in the initial population. One notes that the scores shown in figures 5.43 and 5.44 are the maximal performance of every chromosome obtained using Eq. (5.16).

5.2.8. Resulting Pareto-optimal designs for the modified task

Figure 5.45 shows the Pareto optimal solutions for the modified design task. In the figure four designs are highlighted and the corresponding scenes are shown in figures 5.46 – 5.49. The performance of the design objects that characterize these four designs are shown in figure 5.50.

From the scenes it is seen that the four Pareto optimal designs have certain similar features, which are the following. In this modified design task the stairs take a more prominent place compared to the results in the previous case. That is, the perception of the stairs is higher in the present case compared to the previous one. We also note that the mezzanine is becoming more noticeable as well, which reduces its performance in the present case, since the requirement on mezzanine remained unchanged compared to the earlier design task.

Comparing design D1 with D4 we note that the solutions are quite different regarding the parameter values of the objects. In particular we note that the mezzanine is located more in front of the avatar in D4 reducing the mezzanine's performance in D4 versus D1. Also the orientation of the stairs is different. However, we note from figure 5.50 that the performance scores of the objects in D1 and D2 are quite similar. This is explained considering that the performance value of an object stems from a number of factors that are taken into account together. For instance the degree of perception of the stairs depends on its location in the field of view of the observer, and also on its orientation. Therefore the same perception degree, and thus the same degree of satisfaction for the requirement, may be obtained for different configurations of stairs, which is the case for D1 and D4.

From this example it is noted that for the same performance score of an object the objects' parameters may be quite different in general. Whereas in the opposite situation, i.e. for solutions that have different performance values, it is unexpected that their

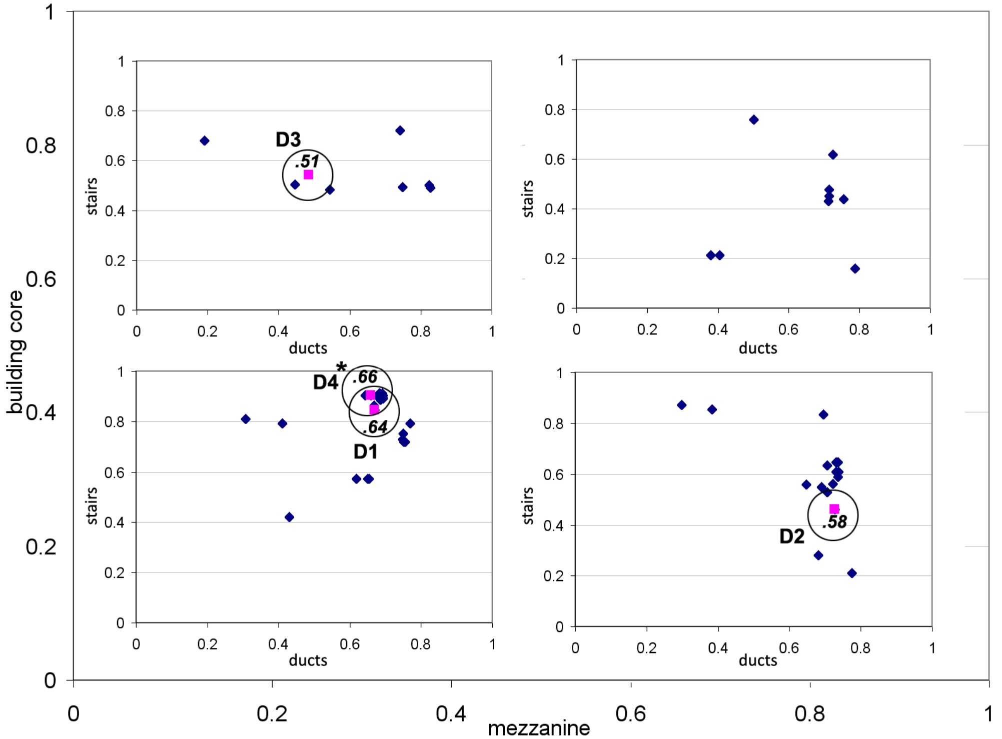


Fig. 5.45 Pareto optimal designs for the modified task

parameters are similar. In other words similarity among solutions in parameter space not necessarily corresponds to similarity in objective space and visa versa. This is seen comparing design D2 with D3. In this case the parameters of the objects are quite similar, however the performance of the design objects differ. This counts in particular for the performance of the ducts object. D2 scores 0.73 in this respect and D3 only 0.48. This is explained considering that in D2 the ducts are made from a darker material, reducing their

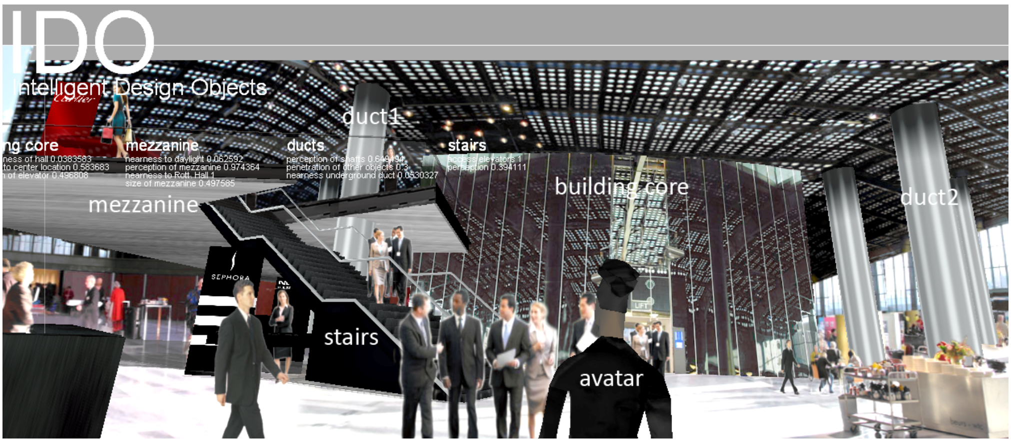


Fig. 5.46 Design *D1*

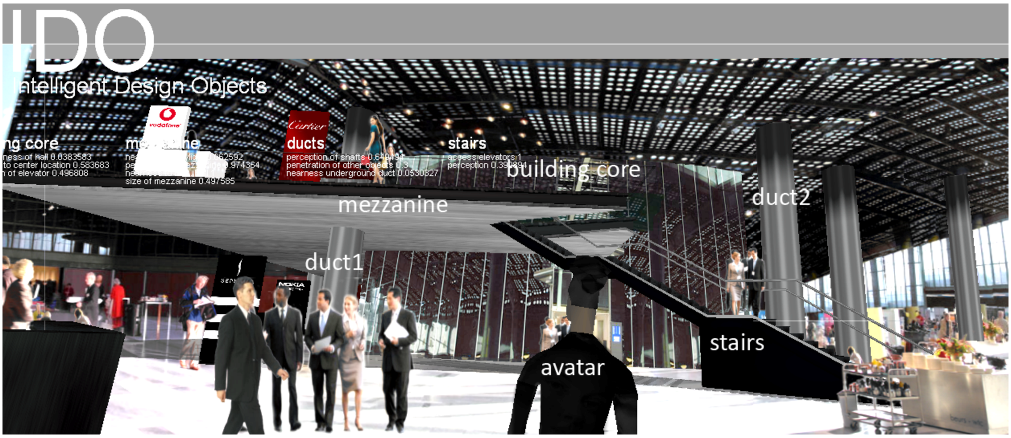


Fig. 5.47 Design *D2*

contrast with the background. This entails lower perception of the ducts compared to D3, which is desirable according to the requirements.

Comparing designs D2 and D3 with D1 and D4, we note that the building core is located nearer to the avatar in D1 and D4 compared to the other two designs. The consequence is that the core's performance is reduced in D1 and D4, due to reduced spaciousness of the hall. That is, D1 and D4 are similar in this respect, while D2 and D4 are also similar. Comparing the designs with regards to their maximal performance score obtained using Eq. (5.16) it is noted that D1 and D4 outperform D2 and D3. Namely D4 has the highest score being .66, followed by D1 with .64, whereas D2 has .58 and D3 scores lowest with .51. This means that in the present task it is preferable from an impartial viewpoint, i.e. without preferring any design object over another, to sacrifice performance of the building core and mezzanine to gain ducts and in particular stairs performance. It is emphasized that this insight is due to the computational cognition of the Intelligent Object approach. It is also clear that according to this cognition design D4 is the most prominent choice among the four Pareto optimal solutions for building execution or as starting point for further elaboration.

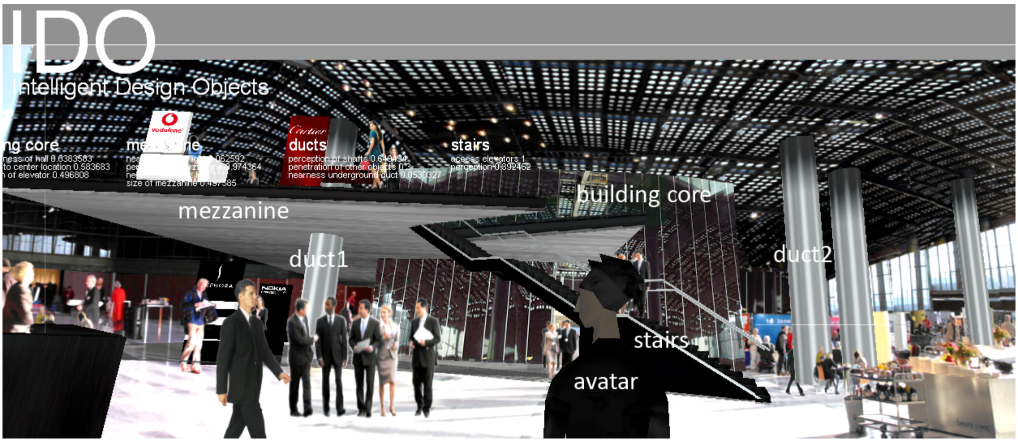


Fig. 5.48 Design *D3*

The detailed differences between a random design from the first generation and Pareto optimal design D4 are shown in table 5.15 for comparison. From the table we note that the Pareto optimal design outperforms the random design regarding most requirements. It is interesting to compare the Pareto front of the present design with the front from the previous design. In the present case the Pareto front is much less concentrated in the same region of the mezzanine-building core space. The reason for this stems from the fact that in the present design the modified requirement for the stairs causes more drastic conflicts among the requirements. In particular the conflict between accomplishment of spaciousness for the hall and the perceptual prominence of the stairs is problematic. To reconcile both demands the Pareto search explores a larger front in the objective space. That is, Pareto solutions span from solutions, where the building core performs well, the stairs performs moderately, to solutions where the building core is moderate and the stairs perform well. Whereas earlier most solutions where found where the building core was moderate and the stair performed well. This is because previously the requirements on the stairs did not favour the other combination as dominant solutions. The previous design can be described as more ‘clear’, i.e. the front was more compact and smooth, whereas the new situation is more ‘tense’, yielding a more scattered front. The scattered Pareto solutions in the second example entail that a designer comparing the solutions of the Pareto front will experience relatively strong differences among designs when he/she is changing his/her preferences only slightly. In contrast to this, in the first exercise a change does not cause strong differences in objective space.

Comparing the first design task with the second one it is emphasized that both are different with respect to their intended purpose. In the first design the mezzanine is intended as an occasional meeting/lunch room, whereas in the second design the mezzanine is intended for retail purposes, where public access is desirable. Therefore it is not relevant to consider the solutions found in the first exercise to be better compared to the second. We can merely discuss the fit for the intended purpose of the criteria in either case. Regarding the conflict between the demanded spaciousness and a prominent location for the stairs, one may wonder why the computational system did not generate a solution where the building core is pushed into the space as far as possible, and the mezzanine is simply made very large. In this way one may deem to accommodate both the spaciousness

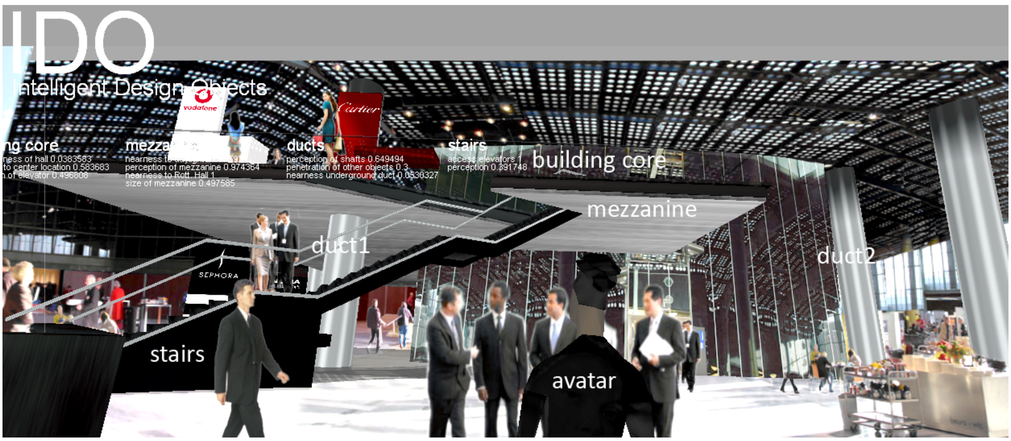


Fig. 5.49 Design *D4*

Table 5.16 Comparison of an initial and Pareto optimal design D4

Node	Output	Initial design	Pareto-optimal design
Design performance	O_4	.29	.66
Performance Building core	$O_3(1)$	.29	.33
Performance Mezzanine	$O_3(2)$	.36	.40
Performance Ducts	$O_3(3)$	.75	.66
Performance Stairs	$O_3(3)$	.15	.90
Functionality of building core	$O_2(1)$	.02	.01
Perception of mezzanine	$O_2(2)$	.45	.40
Functionality of mezzanine	$O_2(3)$	.24	.39
Spaciousness entrance hall	w_1	.29	.39
Centrality core	w_2	.86	.95
Elevator perception	w_3	.19	.14
Nearness mezzanine from daylight	w_4	.15	.22
Perception mezzanine	w_5	.33	.20
Nearness mezzanine from "Rotterdam Hall"	w_6	.99	.97
Spaciousness mezzanine	w_7	.51	.61
Perception ducts	w_8	.76	.62
Penetration of objects by ducts	w_9	.30	.30
Nearness ducts from underground hub	w_{10}	.99	.91
Unobstructed access elevators from entrance	w_{11}	.99	.60
Perception of stairs	w_{12}	.09	.93

requirements and the requirements for prominent stairs at the same time. Although apparently the solution is desirable it has problematic features. One issue is that increasing the size of the mezzanine indefinitely does not increase its performance proportionally. According to the requirements there is an ideal size for the mezzanine, beyond which the satisfaction level is reduced. This is because the additional costs for the mezzanine do not outweigh the gained space, so that this reduces the suitability of the solution. A second issue is that in the proposed situation the elevator access becomes hardly noticeable, and the role of the building core to function as a central hub is impeded.

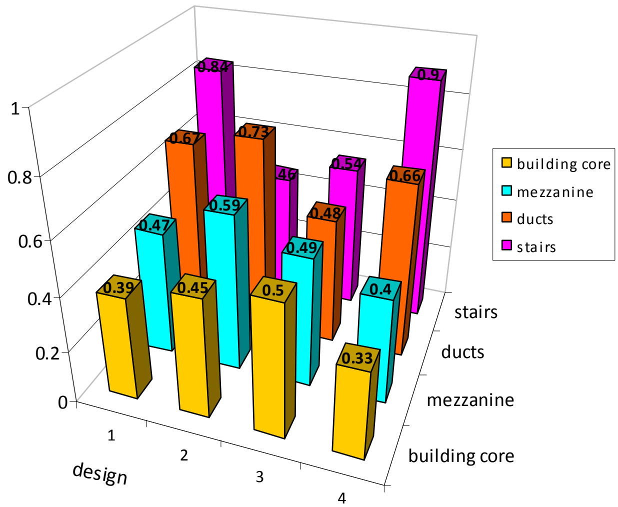


Fig. 5.50 Comparison among designs D1-D4 with respect to their performance values

5.3. Discussion

5.3.1. Applying multi-objective design generation

In architectural design conventionally we arrive at a solution after a complex process of reflection and action, being largely unaware of the detailed steps we took and why we took them. Therefore it is generally difficult to consider what the implications are when we modify our design to investigate if there are comparable alternatives. This means in conventional architectural design without advanced computational assistance there is relatively little flexibility remaining at the end of the design process, and changes may be costly. In contrast to this, in the present approach the flexibility is maintained. This is achieved by not merely focusing on a single 'best' solution, but by generating a diverse range of solution alternatives that are all satisfying our criteria outstandingly.

This feature is particularly important for architectural design, because we are initially unaware, how important the major criteria are with respect to each other. For example in the first application, where a neighbourhood is designed, it clearly difficult to pin-point in advance, how important the size of the gardens is with respect to their visual privacy, although we may have a rough idea about it. And in the second application it is difficult to tell in advance how important the performance of the stairs is with respect to the building core. The reason is because we are initially not aware enough about the consequences our demands have on the corresponding solutions.

This issue is handled in the present approach by letting the computational system generate a Pareto front of solution alternatives. Also the system will highlight the solution with maximal performance among the Pareto solutions when any possible priority among the criteria may be freely chosen. Based on this information we are able to exercise our second-order preferences and select among the solutions with great awareness. This means we may delay our decision on the relative importance among the criteria, until we are in a position to inspect the alternative Pareto solutions. For example in the first application we let the intelligent objects establish the suitable solutions. Thereafter we can immediately see the effect of increasing our preference for the privacy versus the garden size through the behaviour of the intelligent objects. This is because the computational system will offer us the particular solution among the Pareto solutions that matches our higher-level demand.

In this sense an essential benefit we have from using the intelligent object approach is that we are able to concentrate on the higher level issues, while the lower level requirements are already taken care of. This means we can use our experience and knowledge with greater precision, and we are certain about suitability of the final result. This increased precision in the decision making is important in order to prevent failure costs, and in order to make effective use of the available knowledge. In particular this is significant due to the complexity of design, where small changes regarding the criteria may have a significant effect on the final result. By means of the computational approach taking care of a number of requirements at the same time, we are certain that a final decision is less likely to yield undesirable designs that cause inconvenience and great effort to repair afterwards. For example in the second application the perception of the ducts/columns is taken into account, so that during the final design stages or even after the design is built, we will not be surprised to find that the ducts are too much visible from the entrance, which may happen if we would not pay enough attention to such issues next to the other relevant issues.

Next to that another significant benefit from the multi-objective approach is that it enables *innovization* [15]. This refers obtaining innovative solutions and design principles through optimization. These principles are common features among the Pareto optimal solutions, i.e. similar values for a design parameter, which can be discovered inspecting the Pareto solutions. These similarities form design principles that were unknown prior to establishing the Pareto optimal front. For example in design task nr. 1 it is unknown a-priori that all optimal solutions have similar positions for houses *H1-H3, H6*, and G_a1-G_a6. In task nr. 2 it is not known beforehand that the position of the building core is about the same for all optimal solutions and that the mezzanine should preferably be located on the left hand side from the avatar. This way a designer gains deeper insight into the design at hand. This means he is able to simplify the design task by fixing the innovized parameter beforehand, thus freeing computational power for more detailed exploration of the Pareto front.

Moreover it is emphasized that by analyzing the Pareto optimal solutions with respect to their respective maximal performance score may yield further insight into the role design parameters play in the task. Specifically, noting relations between maximal performance and specific settings of design parameters reveals to a designer what aspects of a design are preferably maintained while others may be more readily omitted. It is emphasized that this information is obtained based on the outcomes obtained using computational cognition in the form of Eq. (4.21). This insight could not have been obtained otherwise, and did not exist prior to the execution of the computational process.

5.3.2. Applying the probabilistic perception model

It is emphasized that in the present approach soft criteria are taken into account. This means criteria have an associated tolerance for their partial satisfaction. Examples are the visual perception-based requirements in the applications, where a perception model is used to assess the perceptual properties of the scene.

To illustrate the merits of the perception model let us consider the design of the residential area in the first application, where we desire that the houses of the area have a high visual privacy. When we contemplate a certain arrangement of houses as a solution we need to assess for every garden and living room, how strongly one visually perceives them from the other houses. This information is provided by the perception model pursued. In this way it is avoided that an architect has to spend time to assess the various degrees of privacy of the gardens in the design ‘manually’, for instance through visual inspection of renderings, or assuming multiple viewpoints in a VR environment. But more importantly, in the computational perception case assessments of visual perception are ensured to be made based on the same metric every time. This means the imprecision of the result is minimal. In contrast to this a human observer may not be able to compare the perception form several viewpoints with such consistency, because of the abundance of visual information for each act of observation.

5.3.3. Applying the performance evaluation model

The perceptual properties of the scene are analyzed using the probabilistic perception model. This information is interpreted at the input of the performance evaluation model. There the degree of perception is converted to a degree of satisfaction of an associated requirement. For example the degree of privacy measured by the model is checked against the ideally desired privacy degree using a fuzzy membership function. From the shape of the fuzzy membership functions we note that there is a range where the degree of privacy is deemed high enough, while the satisfaction diminishes as the privacy becomes less. In the same way requirements for a large south garden are also fuzzy. Namely the larger the garden is the better it is, however there is a gradual reduction of satisfaction as the garden becomes smaller.

To deal with such fuzzy requirements, in particular to deal with a number of them at the same time and to reconcile them is a challenging issue. This is because during design we should bear in mind all tolerances associated with every requirement. This way we can evaluate the differences among alternatives precisely. This is clearly problematic due to the large amount of information contained in such soft requirements. As consequence we are not certain that we reached a design that suits our criteria. Employing the cognitive approach presented in this thesis we are able to accomplish this goal being certain that the requirements are duly taken into account altogether. That is, the tolerance for partial satisfaction of our requirements is exploited maximally.

The transparency of the method allows that several experts can combine their criteria in one model. For example in the second design application presented it is simulated that an architect cooperates with an installation advisor. The architect puts forward the requirements on the perception of the ducts, and the installation advisor puts forward the requirement on the distance between the ducts and the underground outlet. Both are taken into account together by means of the fuzzy logic operations. The logic operations using the Gaussian have *local* modelling properties. This means, the output of the computations at an inner node does not directly influence the total output of the model. Therefore the fuzzy neural tree allows increasing the complexity of the evaluation model without limitations imposed by the computational method. This means more design aspects can be taken into account conveniently in the evaluation model. Such aspects may range from functionality, sustainability, costs, cultural value etc. to any other aspect we can think of, provided we have knowledge regarding these issues. The integration of such diverse issues is conveniently accomplished in the present approach. This is because any value measured about a building, say its predicted energy use, can be interpreted as a degree of satisfaction of a corresponding requirement, e.g. a requirement demanding a low energy use. This is done by means of fuzzification at the terminal nodes of the neural tree. Thereafter any degree of satisfaction is processed in the framework of the fuzzy neural tree concept, where the nodes have local properties due to their Gaussian membership functions. Therefore very large number of requirements can be treated together. It is noted that crisp requirements, such as limits imposed by a legal regulation for example, are treated not by means of the fuzzy logic approach. They should be handled using classic bivalent logic. It is noted that due to the transparency of the neural tree structure the results from the bivalent logic can be used in combination with the soft computing approach.

The extensibility of the performance evaluation model has an important implication for cooperative design: An individual person is usually not familiar with the diverse issues playing a role in the different expert domains involved. Therefore a decision maker lacks a solid base to take a final decision with great confidence. This is because he/she is usually unable to verify the overall suitability of a solution with precision. Undesirable features of a design may be discovered in late stages of design or even after realization. Correcting them afterwards is usually expensive. By means of the novel computational approach this bottleneck is alleviated.

5.3.4. Application potential of the cognitive approach

The cognitive approach is generic. That is, the same approach is able to handle different kinds of design criteria. Also the intelligent object concept allows different kinds of objects to be subject to computational design taking into account several soft requirements. Therefore the cognitive approach is applicable throughout the various phases of the building process, ranging from initiative to demolition phase.

For example during the initial design phase the approach is applicable for the design of the shape of a building and its functional organisation. In this case we consider the spaces of the building as intelligent objects. We may aim to achieve low energy use, high functional flexibility, while taking into account the conditions in the surrounding for instance. In later stages we can apply the approach for the design of a detail of a building. In this case the components of the detail are intelligent objects. We may aim to satisfy requirements on energy transmission, wind-density, durability and perception aspects. Or it may be applied during the demolition phase. In this case the decisions concerning the suitable demolition strategy and technology can be seen as an intelligent design object. The aim may be to satisfy safety and efficiency criteria.

Concerning these examples, although they appear in different phases during the design process, there are similarities among the tasks. These are responsible for the fact that a single computational approach is expected to be applicable to all of them. The similarities are that the tasks involve multiple objectives that are conflicting to some extent, and the criteria are usually not crisp but they involve some tolerance for partial satisfaction. For example a demand for ‘low energy use’ does not refer to a single ideal value. It describes a certain bandwidth of acceptable energy consumption, and generally a decision maker’s degree of satisfaction does not immediately vanish in case our ideal consumption value is exceeded. Exploiting these tolerances becomes a major issue as the pressure to use recourses carefully increases. This softness of the requirements in architectural design makes the intelligent object approach uniquely able to reach this goal.

5.3.5. Required time for implementation

The implementation of the cognitive approach described for design applications requires establishment of two essential components. One is the evaluation model, and the second is the search algorithm. As to the first component it is noted that, depending on the type and complexity of the criteria involved, modelling may take a few days up to several weeks. With respect to the genetic algorithm, depending on the amount of parameters subject to genetic optimization, this requires about the same or somewhat more time as the previous component. The latter process may be considered comparable with setting up what is known as a *parametric model* of a design, with the additional effort to establish the parameter identification through genetic algorithm. It is noted that the programming time required for future applications reduces significantly once the cognitive system is set up, depending on the complexity of the task. This is because some program components previously established may be reused. For effective and efficient application of the approach, it is of course necessary that the individual in charge of the application is familiar with the concepts presented in this work and has adequate programming abilities on appropriate platforms. Compared to conventional design techniques without advanced computation it is emphasized that the approach using computational intelligence has important advantages. Namely it is able to search a much greater amount of possible options than a human designer can do consciously, while assessing these options against a more comprehensive and precise representation of design objectives than a human can hold in his/her consciousness. Therefore the additional effort of modelling requirements and establishing a genetic search process is particularly suitable when a high design quality is pursued. It is emphasized that in case the design problem is similar to a previous case already treated, the required time is drastically reduced, so that a superior design quality in terms of assured maximal design performance is attained at minimal cost and effort. For instance setting up the modified design task in application number two took the author merely a few hours.

5.4. Conclusions

Two applications of the computational design system termed *Intelligent Design Objects* are presented. In these applications the system generates solutions that are Pareto optimal with respect to the criteria put forward. It satisfies soft criteria, including visual perception-based criteria. The requirements have a tolerance for partial satisfaction. This means the method's suitability for architectural design is demonstrated where requirements are conflicting, so that they may not be fulfilled totally at the same time. The approach is uniquely able to exploit the tolerance for partial satisfaction or requirements maximizing the overall performance of a design. In this respect computation enables us to become more certain that our design fulfils the criteria. Due to the complexity of design criteria we are generally unable to specify the relative importance among the major criteria of our design in advance. For example we are unable to tell how important our demand for visual privacy is compared to the functionality of a space. This is because it is difficult to foresee the implications our demands have for a corresponding solution. The novel cognitive approach permits us to postpone this commitment until we are able to experience the implications of our criteria for the design. This is realized by generating a collection of Pareto optimal solutions. The solutions provide a variety of alternatives that are all suiting our criteria, while every solution has a different trade-off. Based on these solution alternatives a designer is able to exercise his/her preferences, selecting among the alternatives with great awareness. This is because a designer is able to focus his/her attention on the relative importance among criteria, being freed from the necessity to keep the detailed relations among the design objects in mind at the same time. This is the essential contribution of the computational design, allowing us to handle great complexity with confidence. The application examples presented differ in terms of their respective architectural contexts. This indicates the generic character of the approach and its large application potential. In particular it is noted that the special type of computations performed in the performance evaluation model allow integrating a large number of different requirements. These may range from sustainability, functionality aspects, comfort, and safety aspects etc. Due to its generic character the cognitive approach is applicable throughout the various phases of the building process, ranging from initiative to demolition phase.

References

1. Ciftcioglu, Ö., Bittermann, M.S.: Solution diversity in multi-objective optimization: A study in virtual reality. Wold Congress Comp. Intelligence, Hong Kong (2008)
2. Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, Ö.: Performance-based Pareto optimal design. In: Rusák, I.H.a.Z. (ed.): TMCE 2008, Izmir, Turkey (2008)
3. Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, Ö.: Multi-objective optimization in the construction industry. AEC 2008, Antalya, Turkey (2008)
4. Ciftcioglu, O., Bittermann, M.S., Sariyildiz, I.S.: Building performance analysis supported by GA. In: Tan, K.C., Xu, J.X. (eds.): 2007 IEEE Congress on Evolutionary Computation. IEEE, Singapore (2007) 489-495
5. Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: A neural fuzzy system for soft computing. NAFIPS 2007. IEEE, San Diego, USA (2007) 489-495
6. Bittermann, M.S., Ciftcioglu, O.: A cognitive system based on fuzzy information processing and multi-objective evolutionary algorithm. IEEE Conference on Evolutionary Computation - CEC 2009. IEEE, Trondheim, Norway (2009)
7. Gibson, J.J.: The Ecological Approach to Visual Perception. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey (1986)
8. Foster, J.: The Nature of Perception. Oxford University Press (2000)
9. Palmer, S.E.: Vision Science. MIT Press, Cambridge (1999)
10. Levine, M.W., Sheffner, J.M.: Fundamentals of Sensation and Perception. Addison-Wesley, London (1981)
11. Sato, H., Aguirre, H.E., Tanaka, K.: Controlling dominance area of solutions and its impact on the performance of MOEAs. Springer, Heidelberg (2007)
12. Laumanns, M., Thiele, L., Deb, K., Zitzler, E.: Combining convergence and diversity in evolutionary multiobjective optimization Evolutionary Computation 10 (2002) 262-282
13. Goldberg, D.E.: Genetic Algorithms. Addison Wesley, Reading, Massachusetts (1989)
14. Deb, K., Goldberg, D.E.: A comparative analysis of selection schemes used in genetic algorithms. First Workshop on Foundations of Genetic Algorithms. Morgan Kaufmann, Bloomington Campus, Indiana, USA (1990)
15. Deb, K., Srinivasan, A.: Innovization: Innovating design principles through optimization. GECCO '06, Seattle, Washington (2006) 1629-1636
16. Deb, K.: Multiobjective Optimization using Evolutionary Algorithms. John Wiley & Sons (2001)
17. Miettinen, K.: Nonlinear Multiobjective Optimization. Kluwer Academic, Boston (1999)

6. Conclusions and recommendations

6.1. Summary

A novel computational system for architectural design is presented. It generates designs that satisfy the criteria put forward by decision makers, such as architects.

Architectural design requires knowledge spanning alpha, beta, and gamma sciences. This diversity is needed because buildings address human needs that range from basic needs for comfort and safety, to higher level needs, such as needs for aesthetically appeal. The higher level needs largely determine the architectural and cultural value. They are generally satisfied via the form quality of a building. Therefore design criteria usually include visual perception aspects, which are difficult to grasp and handle with precision. To alleviate this issue as first component of the approach a visual perception theory is developed. Based on the theory perceptual properties of spaces are evaluated with comparable precision, as this is conventionally done for other aspects of designs, such as structural properties, sustainability aspects etc. In this way subtle perceptual issues can be weighed against other criteria, ensuring that sacrifices of performance aspects are minimized in accordance with the preferences.

As design requirements are generally conflicting, for example high perceptual value versus low cost, they are usually not satisfied totally at the same time. Therefore in architectural design we tolerate partial satisfaction of requirements to some extent. This means design requirements are generally *soft*. This softness makes it difficult to compare alternative designs. In particular it is challenging to bear in mind the tolerances associated with every requirement, while taking multiple requirements into account at the same time. This problem becomes apparent in the common difficulty to tell whether a design becomes better or worse when we modify a design object. For example when we change the position of a column this may be positive from a structural viewpoint. However the same act may have negative effects from the perception viewpoint, and it is difficult to assess the impact on the overall design.

In this thesis this problem is tackled using a novel computational method that is able to deal with the softness of design requirements. Using this method, perceptual requirements are considered together with other requirements in a model. The respective tolerances put forward by an architect are taken into account. This way the performance of a design is evaluated in an integral manner.

Evaluating the performance of a design based on the criteria is not enough to ensure that suitable designs are found. The difficulty is that usually we should compare a multitude of solution options to be confident we reached a suitable solution. That is, an excessive amount of combinations among design parameters should be considered. However, this is generally not possible due to time limitations. This problem is alleviated in the present approach by letting an algorithm explore possible solutions in a goal-oriented manner. The algorithm uses the output from the performance evaluation model to generate solutions that match the given criteria.

Due to the complexity of criteria in architectural design we are generally unable to express in advance the relative importance among criteria. For example it is problematic to tell how important the perception aspects of a design should be taken compared to its functional performance. Both are very important but are the equally important? This is difficult to answer a-priori because it is unknown in which way precisely both requirements are conflicting, and what impact the conflicts will have on solutions. Therefore it is inappropriate to commit in advance regarding the importance among criteria and to search for a single optimal solution. The computational system developed in this thesis allows postponing this commitment, because it generates a *range* of alternative solutions. All of them match the criteria, while every one of them has different strengths and weaknesses for different criteria. These solutions are known as Pareto-optimal solutions. It is a matter of cognition to select among these solutions, since they are equivalently suitable in some sense. Interestingly, due to the particularity of the present approach the computer is able to distinguish among these solutions, pin-pointing a solution that is most potent for the intended purpose of the design. This is accomplished in the present approach by letting the computer identify a certain Pareto-optimal solution; namely the solution having the highest performance when preferences regarding the criteria may be freely selected.

Based on this solution and the other Pareto optimal ones, we can explore the trade-offs among the solution alternatives. This way we are able to make more confident statements on the importance among the criteria. This is because we can directly observe the implications our preferences have on the solutions. During this exploration we do not need to bear the multitude relations among design objects in mind, because they have already been taken into account through the computations. This means we are able to focus on the interplay among complex criteria rather than having to ensure the satisfaction of elemental requirements at the same time. Therefore we are able to exercise our preferences with great awareness. As a metaphor we can say we are conducting an orchestra focusing on the higher-level aspects of the peace, in place of being busy to play every instrument at the same time. Based on the Pareto optimal solutions a designer gains insight into his design in the sense that he/she realizes the role of certain solution parameters more clearly. This way he/she is able to adjust design criteria consciously, so that they match closer to his/her ultimate preferences, and then let the computational process run again based on the improved criteria. It is expected that this will be necessary several times during a design process. It is emphasized that this process is an interactive process where machine and human cognition are interfaced via the intelligent objects. This interactive process may therefore be termed *Interactive Intelligent Design Objects* (I^2DO). It is to be noted that this approach does not claim full automation of the design process, as it involves a decision maker in the loop for interactivity.

The intelligent objects are applied in two architectural design applications that differ with respect to scale and context. This demonstrates the effectiveness and generic character of the cognitive approach.

Exercising second-order preferences during a human machine interaction, where multiple, soft criteria are treated by means of advanced computations, makes the approach a *cognitive* approach. It is noted that the treatment of design as a cognitive process, where the Pareto concept plays a central role, bears potential to function as an architectural formalism.

6.2. Novelties

The approach developed is based on understanding architectural design as a cognitive process, where solutions are generated that match given criteria, while the designer remains free to exercise his/her second-order preferences. This is a novel approach. Its development required tackling several state-of-the-art scientific challenges.

A major issue concerns dealing with the softness of criteria. Architectural design criteria are generally soft, because they contain conflicting requirements, some of which refer to elusive phenomena, such as visual perception. Soft criteria are conventionally considered incomputable. The major accomplishment of the present approach is treatment of soft criteria by means of computation.

A bottleneck in this respect is to handle visual perception aspects with precision. These aspects are significant in architectural design, because they greatly determine the value of a building. Despite this significance the perception phenomenon is merely vaguely understood up till now. This is because the perception process is very complex: We see in the *brain* and not merely with our eyes, i.e. our visual *attention* determines to what extent we are aware of objects around us. Based on these considerations in the present thesis a theory of visual perception is presented. In this theory visual attention and perception are modelled using probabilistic terms. The probabilistic approach is entirely novel, noting that existing approaches were either statistical or deterministic. Due to the complexity of the eye-brain process the existing approaches were unable to describe the general relation between an environmental stimulus and a corresponding seeing event. In contrast to this, the present probabilistic approach describes this very relation. The model is generic, i.e. it is not restricted for any particular spatial situation or for any particular observer. Based on the model the perception properties of spaces are analyzed with precision. This way, subtle perceptual differences can be taken into account in the design.

A second issue dealing with the softness of criteria concerns the fact that many requirements are not sharply defined. Examples of such fuzzy requirements are demands for *high visual openness*, or *high visual privacy*. It is noted that beyond perceptual requirements, many other design requirements involve imprecision. Examples are demands for *low energy use* or *high functionality*. The fuzziness of the requirements refers that they have an associated tolerance, where partial fulfilment of a requirement is still deemed acceptable to some extend. Dealing with the fuzziness of requirements forms a bottleneck for computational design methods. This is because computation conventionally requires sharply defined input information. This is the reason why the methods from the domain of classical artificial intelligence are unable to handle architectural design requirements.

Related to the problem of fuzziness of requirements is the problem that design criteria consist of a number of aspects and sub-aspects. When we demand a highly functional building, the high functionality may depend on high accessibility, and adequate size and proportion of the spaces. Accessibility again depends on several sub-issues and so on. The multiple attribute relations involved in design criteria pose an additional difficulty to assess the performance of designs, next to the fuzziness of requirements.

In the present approach the fuzzy nature of requirements and the multiple attribute relations involved in the criteria are handled. This is accomplished employing a novel modelling approach. It is based on fuzzy logic processing executed within a special neural structure. The fuzzy information processing allows dealing with the fuzziness of design requirements, and the neural structure handles the attribute relations. This symbiotic approach is uniquely able to address both issues at the same time, while it maintains the transparency in the model. This means an architect is able to insert his knowledge regarding the design criteria into the model, forming an ontology of the design.

Having established means to evaluate the performance of designs including soft criteria, a consequent challenge is to arrive at designs that satisfy the criteria. In this respect generally many diverse potential solutions should be investigated. This poses problems for conventional computational approaches. Another problem of conventional approaches is that they generally require specification of the relative importance among the criteria. This is usually problematic in the beginning of a design, because the consequences of such commitment are unknown. These problems are alleviated in the present approach by employing evolutionary search for multi-objective optimization. The method generates a large amount of solutions that match the criteria Pareto optimally, instead of reaching merely a single 'best' solution. An architect is able to explore these designs afterwards, exercising his/her preferences during the selection. A challenging issue in this respect is to generate a diverse range of solutions, so that we gain a broad insight into the potential solution alternatives. The issue of diversity among Pareto solutions is a state-of-the-art concern in evolutionary computation. In the present work this problem is alleviated by employing a novel concept to grade the potential solutions during the evolutionary search process. Compared to conventional approaches the novel concept enhances the effectiveness of multi-objective search increasing the diversity of solutions on the Pareto front. Another novel feature is that the computer is able to distinguish among the Pareto optimal solutions with respect to their maximal performance. This is an act to be classified as *computational cognition*, as it includes determining second-order preferences that were not known before executing the design.

Summarizing, the scientific novelties developed in this thesis are:

- Performance evaluation: A method to deal with multi-facetted, soft criteria is developed
- Design generation: Multi-objective evolutionary algorithms are enhanced by a novel ranking concept, yielding diverse solutions that match the criteria
- Perception modelling: A probabilistic theory of visual perception is founded, allowing to treat the perception properties of designs with precision
- The above components are integrated forming a unique intelligent system for design with cognitive features.

6.3. Expected practical benefits

The complexity of designs is more and more increasing. Performance criteria are diverse spanning several domains of expertise. They range from technical, functional and environmental aspects, to perceptual and cultural aspects. Many criteria are conflicting. The amount of possibilities that need investigation to satisfy the criteria is enormous. At the same time clients wish to be confident that the design they are going to build is matching their preferences. They demand some evidence that no designs were overlooked that would be more suitable. Therefore, advising a forward-thinking client in such a way that he is confident he got the best solution becomes more and more challenging. In particular the reasoning behind preferred solution directions should be comprehensive, consistent and transparent. The difficulty is due to involvement of diverse, linguistic concepts in the decision criteria, such as *operational certainty*, *financial certainty*, *sustainability*, or *attractiveness*. This makes differentiation among solutions imprecise, and hampers communication of the reasoning between experts and clients. On top of that, the most suitable solution is to be identified in ever shorter periods of time. These issues pose sever challenges for decision-makers, architects and advisors. The novel computational approach presented in this thesis alleviates some of these issues as follows.

A major issue is the evaluation of designs taking into account the diversity and complexity of the criteria. Judging a design needs consideration of many facts at the same time. Existing analysis tools, such as computational fluid dynamics (CFD) and finite element methods (FEM) together with expert assessments provide elemental information about a design. The expert assessments include judgements of aesthetic aspects in immersive virtual reality (VR) environments, for example. However, when design decisions are taken it is necessary to interpret and combine the results from the tools. This way the suitability of solutions is judged in a holistic manner. Improving this judgement processes is especially proficient in the early design stage, where uncertainty is high and the influence on later stages is crucial. The novel method for performance evaluation allows decision makers to integrate their diverse requirements into one model, while the requirements may be quite different in nature. For example functional and perceptual aspects are treated together. Essentially the computational model ensures that a decision maker can concentrate on several detailed aspects one by one, while the mode ensures that they are considered consistently altogether. This way performance evaluation achieves a higher quality compared to the conventional case. Due to the transparency of the performance evaluation model and the Pareto concept involved, the rationale behind the solutions is transparent. This way a client commissioning a design is able to gain insight into the potential trade-offs he/she is about to accept.

Keeping the interdependencies among design attributes in mind while designing is an effort consuming attention. This is alleviated using the cognitive approach presented. The approach makes use of the performance evaluation model during design generation. This means the computer is able to search for suitable designs based on 'awareness' of the entire range of criteria. Next to this, computation is able to investigate a large number of alternative designs in a short time. This is not feasible otherwise.

As a computational solution set has been built, alternate designs are explored by varying the parameters. The generative system can handle effectively up to five objectives, and has no restriction regarding the number of variables playing role on the objectives. The amount of variables characterizing a solution is only limited by available computational time and power.

The strength of the approach is that project solutions can be assessed without any presupposition, and confidence of finding the best solution is increased.

It is noted that the computational design process is in a symbiotic partnership with the designer. That is, it takes care of those aspects of designing which are computationally intensive, letting a designer exercise his/her creative ideas with greater effectiveness. In the initial phase of design there is uncertainty regarding the relative importance of design criteria. This means we are essentially charged with *two* tasks during design. We should arrive at designs that satisfy our criteria, *and* we should form a better understanding of the significance of the criteria as we are progressing in the design. The multi-objective search takes care of the former issue, so that we can focus on the importance among criteria, being sure that our criteria are already taken into account. Based on the repertoire of solution alternatives generated, the intelligent design objects respond to our higher order preferences, showing us the solution that fits both our criteria and our second order preference. This means through intelligent objects a designer is controlling the design in a more sophisticated manner than if he/she were to manipulate the design objects in conventional ways.

It is emphasized that the approach facilitates taking care of diverse criteria at the same time. It is expected that this facility may have a particularly boosting effect on *cooperative* design. This feature is relevant in the face of increasing diversity of criteria in architectural design. This trend is also noticed by the founders of the cradle to cradle concept. They consider the problem how to reconcile the already diverse demands in architectural design, with aspects of equity, ecology and economy. One source hampering the treatment of such diversity in cooperative design is the fact that the reasoning behind design actions an expert undertakes is not made transparent to other cooperating experts. Therefore usually trade-offs inherent to a certain design action are not apparent in the decision making. This implies decision makers are unable to be very confident with their decision. The reason is that they are lacking information on solution alternatives and their inherent trade-offs. Due to the transparency of the reasoning in the intelligent object approach, new avenues are provided for cooperative performance-based design.

Applications of the method are diverse due to the generic character of the intelligent object approach. That is, the approach is able to handle different kinds of requirements. Therefore the cognitive approach is applicable throughout the various phases of the building process, ranging from initiative to demolition phase, while the method remains essentially the same. The differences among the various possible applications are that the relevant criteria are different, and the design objects are of different kind. For example during the initial design phase the approach is applicable for the design of the shape of a building and its functional organisation. In this case we consider the spaces of the building as intelligent objects aiming to achieve low energy use, high functional flexibility, while taking into account the conditions in the surroundings for instance. In later stages we can apply the approach for the design of a detail of a building. In this case the components of the detail are intelligent objects, aiming to satisfy requirements on energy transmission, wind-density, durability and perception aspects. Or it may be applied during the demolition phase. In this case the decisions concerning the suitable demolition strategy and technology and their parameters can be modelled as intelligent design objects, where the aim is to satisfy safety and efficiency criteria.

Concerning these examples, although they appear in different phases during the design process, there are similarities among the tasks. These are responsible for the fact that a single computational approach is applicable to all of them. The similarities are that the tasks involve multiple objectives that are conflicting to some extent. Also the criteria are usually not crisp but they involve some tolerance for partial satisfaction. For example low energy use does not refer to a single ideal value, but it describes a certain bandwidth of acceptable energy consumption. And generally our degree of satisfaction does not immediately vanish in case our ideal consumption value is exceeded. Exploiting these tolerances becomes a major issue as the pressure to use recourses carefully increases. Its capability to exploit the softness of the requirements makes the intelligent object approach uniquely able to deal with the size of the problems resulting from the environmental pressures.

In summary, the expected benefits of the cognitive approach are:

- Project solutions can be assessed without presupposition, and confidence of finding the best solution is increased, since solutions are obtained with increased certainty that they match the preferences
- Diverse criteria are treated in perspective, so that an overall performance score is obtained, and alternative designs are compared with precision, while non-technical criteria are taken into account together with technical criteria
- Architects are able to focus on the higher-level aspects of their design, letting the computation take care of the lower-level dependencies among the design objects

6.4. Recommendations

Based on the present work a number of relevant topics become apparent to expand the cognitive approach presented. These are illustrated in the scheme in figure 6.1.

From the scheme it is noted that a major topic concerns the integration of other design criteria into the approach. Beyond the perception assessments, and geometric relations among objects, other aspects of design are subject to measurement, in order to increase the comprehensiveness of the criteria treated in a design. Such issues are for example to predict the airflow in a space and around buildings, to model the structural stress propagation, and to predict the energy loss of the building. This is accomplished using appropriate models. These are based on e.g. computational fluid dynamics (CFD), finite element methods (FEM) and finite volume methods (FVM). Also cellular automata (CA) may be appropriate when complex dynamics of e.g. urban developments need to be modelled. Such models deliver facts about a design, which are subject to interpretation in performance analysis. A number of models are indicated in the figure nearby the encircled letter *a*. For a number of relevant parameters sufficiently accurate models exist. However, due to the need to explore a large number of potential solutions searching for suitable ones, it is a relevant issue to investigate the possibility to reduce the amount of computation required to obtain results from the models. In this respect the gain in speed should be weighed against the loss in accuracy of a model. The latter should be commensurate to what is needed in the particular stage of the design. Also models need to be developed for aspects that are not tackled up till now. Examples of such aspects are measurement of accessibility, or to extend the perception model, so that is includes biased observer conditions, and also takes *gestalt* aspects into account. Also developing a comprehensive privacy concept that includes aspects beyond visual perception is desirable. Together with the newly provided information coming from the measurement models the neural tree should be expanded with requirements and criteria involving these outcomes. Duly integrating such models, the intelligent object behaviour is to become more and more comprehensive.

Concerning the identification of Pareto solutions it is an interesting relevance to investigate the systematic selection of the angle of tolerance in the relaxed dominance concept. This should further enhance the formation of Pareto fronts, which is significant when the amount of criteria increases further.

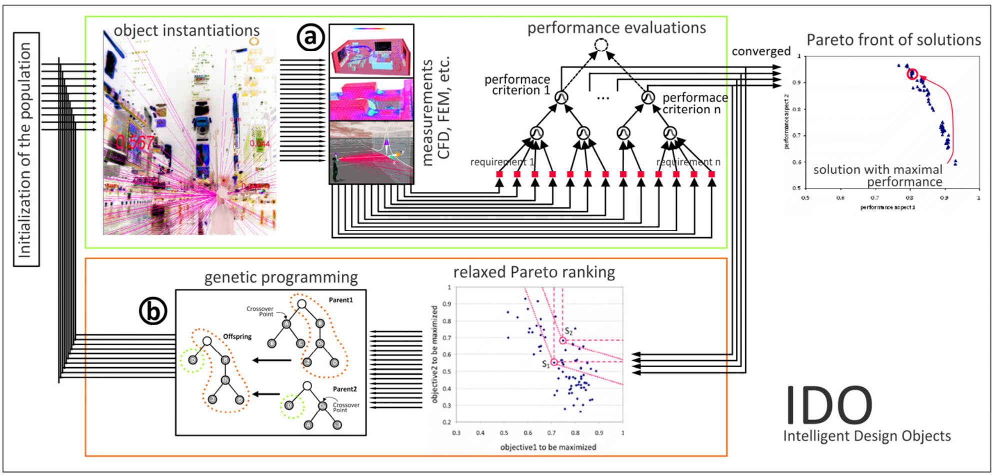


Fig. 6.1 Extensions of the cognitive system based on *Intelligent Design Objects*

Remembering that the intelligent objects are essential components of a cognitive design approach, where a designer is interacting with machine cognition, it is an interesting relevance to observe details of this interaction. This may provide further insight into human creativity. It is noted that in the design generation, a designer has to specify the boundaries for the generated parameter values. In the inception of a design such specification may be problematic to make, since it already requires some rather concrete idea about the solution. In order to boost the design generation component to deal with greater uncertainty regarding the potential solutions, it is significant to enhance the exploration capabilities of the algorithm. It is expected that using a *genetic programming* approach instead of a genetic algorithm may provide significant improvement in this respect. This is shown in the figure at the encircled letter *b*. This expands the range of possible solutions explored by the computations, so that highly unexpected solutions with outstanding performance can be reached. We can say that this extension of the method boosts the 'creativity' of computational design generation. This is particularly significant as one can expect that it will increase the effectiveness of the cognitive approach in applications.

Coupling such rich evolutionary computation methods with the soft performance evaluation proposed in the present thesis it is expected that the computational 'creativity' exhibits more and more features of natural creativity. This may be a breakthrough to deal with the challenging issues of the present time, as this has been already envisioned by Nobel laureate Herbert Simon, when he made the following statement[1]:

"The first industrial revolution showed us how to do most of the world's heavy work with the energy of machines instead of human muscle. The new industrial revolution is showing us how much of the work of human thinking can be done by and in cooperation with intelligent machines. Human minds with computers to aid them are our principal productive resource. Understanding how that resource operates is the main road open to us for becoming a more productive society and a society able to deal with the many complex problems in the world today. … The tools now being forged for aiding architectural design will provide a basis for building tools that can aid in formulating, assessing, and monitoring public energy or environmental policies, or in guiding corporate product and investment strategies."

The rapid changes in our environment may be taken as a reminder of the fact that our intellectual ability to comprehend and interact with the complex relations in our environment is limited. Architecture is intertwined with a number of relevant global issues, including ecological and economic welfare. Therefore there is pressing demand for a new kind of design technology, i.e. decision-making technology allowing us to expand our natural mental limitations.

[1] Simon, H.A.: Decision making and problem solving. Research Briefings 1986. National Academy of Sciences, Washington, DC (1986)

Appendix A. Colour

This Appendix provides details about the colour perception component of the perception model presented in Chapter 2.

Next to geometry, colour plays a significant role in perception, influencing the degree of awareness an observer has for an object. The influence is based on the concept of colour difference, more specifically, *perceptually uniform* colour difference. The concepts of colour difference and perceptual uniformity require an explanation how colour is numerically specified. This is provided in section A.1.

Section A.2 describes the model for colour specification conventionally in use. This is known as the *C.I.E. standard observer* with its associated *C.I.E. XYZ* colour space.

In computer-aided design, colours are often specified in *RGB* colour space. The *RGB* colour space is not perceptually uniform. This means colour difference measured in this space does not represent the normal human judgement of colour difference. Therefore it is necessary to describe the conversion from *RGB* to *C.I.E.* $L^*a^*b.^*$ This way perceptually uniform colour difference can be obtained. This is described in section A.3.

In section A.4 the significance of colour is investigated. This is done by comparing a chromatic scene with its greyscale counterpart. Also the mean colour difference among the basic colours of the visible spectrum is computed, since this value plays a role in the colour perception model presented in this thesis. This is followed by some details on the interior perception analysis in Chapter 2.

A.1. Numerical specification of Colour

Colour is a perceptual effect of light. This means colour exists in human consciousness and not “in nature”. When we say an object “has a colour” the colour we attribute to the object is the content we have in our consciousness. This content we have due to sensory and neural processing of the light the object reflects. The light enters our eye and is processed in the brain, yielding the sensation of colour. It may be interesting to note that when we say a tree has green leaves for example, the green we see is the part of the light the leaf reflects and does not absorb. Apparently the leaves of a tree are not using the energy of the “green” part of the light spectrum.

A.1.1. Light and colour

Light visible for a human has wavelengths that range from 400nm (violet) to 700nm (red). The colour belonging to the different wavelengths of the visible spectrum is illustrated in figure A.1. Light we usually experience contains a mixture of many different wavelengths. This is in contrast to monochromatic light, which contains only one wavelength. Approximately monochromatic light can be produced by using a laser.

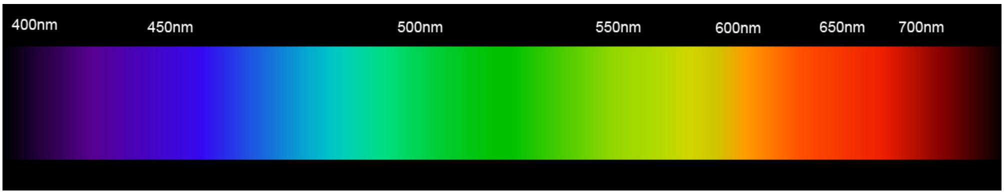


Fig. A.1 The visible spectrum of light

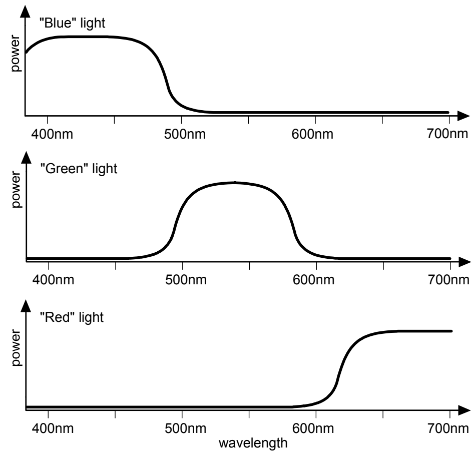


Fig. A.2 Spectral power distribution for a blue, green and red light, as separated by a prism.

Next to wavelength another quantity characterizes light. This is the *radiant energy flux* or *power* at a certain wavelength band. The characteristic composition of light in terms of power with respect to wavelength is termed a spectral power distribution (SPD). An SPD gives the power of a light per unit wavelength at the different frequency bands of the light.

Some exemplary SPDs are shown in figure A.2. The figure shows SPDs for “blue”, “green” and “red” light. The colours are obtained by separating “white” light by a prism. Light we perceive as white is a mixture of all wavelengths throughout the visible spectrum, where the power of the light at any wavelength is similar. Black colour is experienced when no or not enough visible light is impinging on our receptors.

It is interesting to note that for different SPDs a human observer can have an identical colour impression [1]. This phenomenon is referred to as *metamerism*. This phenomenon clearly shows that colour is a perceptual phenomenon. The reason for metamerisms is the constitution of the human retina. The retina is populated by different type of receptor cells that respond differently to light with the same wavelength characteristics. The cells are referred to as cones. The different type of colour receptor cells are respectively referred to as *S*, *M*, or *L* cones. The name refers to the domain in the wavelength spectrum, the receptor type is sensitive to: *S* stands for short, *M* for medium and *L* for long wavelength. Each of the cones has its own response curve to an SPD.

Figure A.3 shows a spectral sensitivity on an average human observer expressed by the photon absorbance of the respective receptor type for a specific wavelength. The values are obtained by means of experiment [2]. It is noted that the response of the *M*-cones and *L*-cones are overlapping to a large extent. Therefore the commonly held idea that the three cones of the human retina are respectively sensitive to blue, green and red is a very lose conception.

A.1.2. C.I.E. standard observer

In vision science and colour science it is relevant to uniquely specify a colour from the SPD of a light stimulus. Due to the perceptual nature of colour this is a challenging issue.

For pragmatic reasons the colour science community considered it desirable to specify colours in terms of the relative intensity of monochromatic primary colours blue, green, and

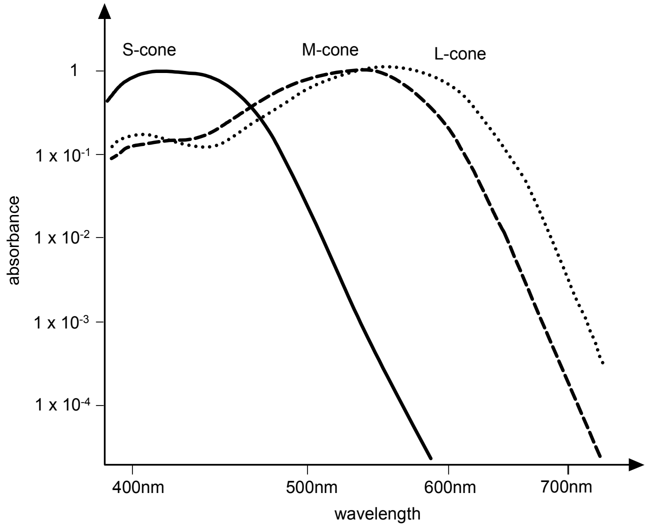


Fig. A.3 Sensitivity of the cones on the human retina with respect to wavelength of light expressed by the respective energy absorbance of the cones; source: [2]

red, rather than in terms of the energy absorbance of the *S, M* and *L* cones. For this purpose appropriate weighting functions are established. This way an SPD can be converted to a colour description by means of three numbers belonging to the primary colours. This is explained below.

In order to establish the weighting functions the following psychophysical matching experiments were conducted [3, 4]. The experiments rely on the law of *trichromatic generalization*, which states that (i) stimuli with same specifications look alike to an observer with normal colour vision under the same observing conditions, (ii) stimuli that look alike have the same specification, and (iii) the numbers comprising the specification belong to continuous functions [5]. In such matching experiments a large group of observers is asked to produce a certain colour by adjusting separately the intensity of three monochromatic primary colours red, green and blue.

The observers should combine the three colours in such a way that the combination matches as close as possible to a given monochromatic test sample with a known wavelength. The colours are typically presented in the two halves of a bipartite visual field. The wavelength of the test sample is varied throughout the visible spectrum. The experiment is repeated at sufficiently small intervals. The resulting intensities are normalized to have constant area beneath them, as given by Eq. (A.1) and plotted against the wavelength of the test sample. The result is shown in figure A.4.

The resulting functions are called the *colour matching functions* for the experiment. This is because they represent the colour matching properties of the observer. They are labelled $\bar{b}_\lambda$ for blue, $\bar{g}_\lambda$ for green, and $\bar{r}_\lambda$ for red in the figure. In this case *λ* at the index refers to fact that the specified amount of the wavelength is a primary colour with the specific wavelength *λ*. In the experiments referred above, the wavelength of blue was taken at 435.8 nm, green at 546.1 nm, and red at 700.0 nm. It is noted that the colour matching functions are dimensionless conversion factors from spectral power distributions to their tristimulus representation [6]. Normalization ensures that [7]

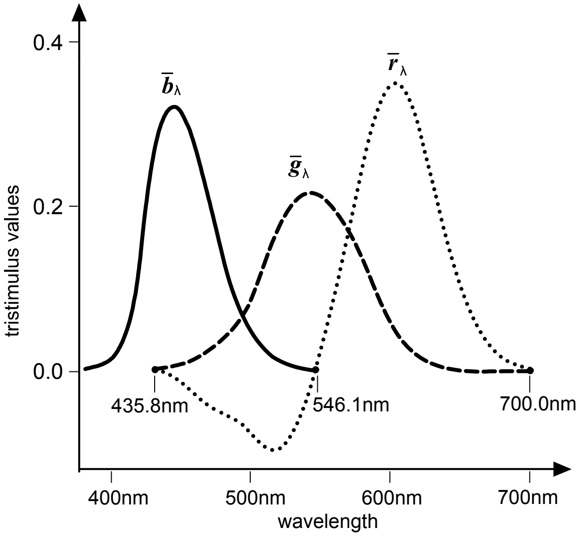


Fig. A.4 Colour matching functions based on *b, g, r* primary stimuli defining the *1931 C.I.E. standard observer*; source: [5]; the units of the primary stimuli for blue, green, and red are normalized, so that the radiant power ratios are 72.1:1.4:1.0

$$\int_0^\infty \bar{b}_\lambda \, d\lambda = \int_0^\infty \bar{g}_\lambda \, d\lambda = \int_0^\infty \bar{r}_\lambda \, d\lambda \qquad \text{(A.1)}$$

so that the energy absorbed by each receptor type will be the same. This normalization is convenient for the later use of the colour matching functions for the specification of an SPD. To achieve the normalization the units of the primary stimuli for blue, green, and red are chosen to have the radiant power ratios 72.1:1.4:1.0.

Not all monochromatic colours of the visible spectrum could be matched using the matching procedure described above. When this was the case, the observer was able to increase the intensity of one of the primaries of the test colour sample, and still achieve a match. For these cases, the amount of the primary added to the test colour was considered to be a negative value at the corresponding matching function of the primary. In this way, for the entire visible spectrum monochromatic light with a specific wavelength can be expressed in terms of three numbers, which express three amounts of the primary colours termed B, G, and R that represent the light. The values for B, G, and R are termed *tristimulus values*. They are obtained by multiplying the SPD with the weighting function of each primary and thereafter integrating the respective function. This is given by Eq. (A.2).

$$\begin{aligned} B &= \int_0^\infty \bar{b}_\lambda(\lambda)\; P(\lambda)\, d\lambda \\ G &= \int_0^\infty \bar{g}_\lambda(\lambda)\; P(\lambda)\, d\lambda \\ R &= \int_0^\infty \bar{r}_\lambda(\lambda)\; P(\lambda)\, d\lambda \end{aligned} \qquad \text{(A.2)}$$

where *P(λ)* is the spectral density of radiant flux per unit wavelength interval specified through the SPD of the light; *λ* is visible wavelength. The dimensions of the radiant flux are those of power, and appropriate units are Watts. The colour matching functions obtained by [3, 4] were adopted by the Commission International de l'Éclairage (C.I.E.) and define the

1931 C.I.E. standard observer. The set of colours that can be described via R, G, and B form a colour space known as the *1931 C.I.E. RGB* space.

A.1.3. C.I.E. (r.g) chromaticity diagram

The tristimulus values can be normalized based on the total strength of primary stimulus as given by Eq. (A.3). This is due to Grassmann's *proportionality law*. The law states that if a colour *A* matches a colour *B*, then *αA* matches *αB*, where *α* is any positive factor by which the radiant power of the colour stimulus is increased or reduced, while its relative SPD is kept the same [8].

$$
\begin{aligned}
b &= \frac{B}{B+G+R} \\
g &= \frac{G}{B+G+R} \\
r &= \frac{R}{B+G+R}
\end{aligned}
\qquad \text{(A.3)}
$$

The resulting values for *b, g,* and *r* are termed chromaticity coordinates. They are convenient to specify a colour. This is because the normalization ensures that *b=1-(r+g)*, so that a colour can be specified using only two numbers, while the third number adds no new information. Therefore the colours can be plotted as shown in figure A.5. The colour located at $b_E=g_E=r_E=1/3$ belongs to the *equal-energy stimulus E*, which is white.

We note that the matching function for red shown in figure A.4 takes negative values in the wavelength range around 500 nm. The *C.I.E.* considered this property undesirable from the viewpoint of computational convenience at that time.

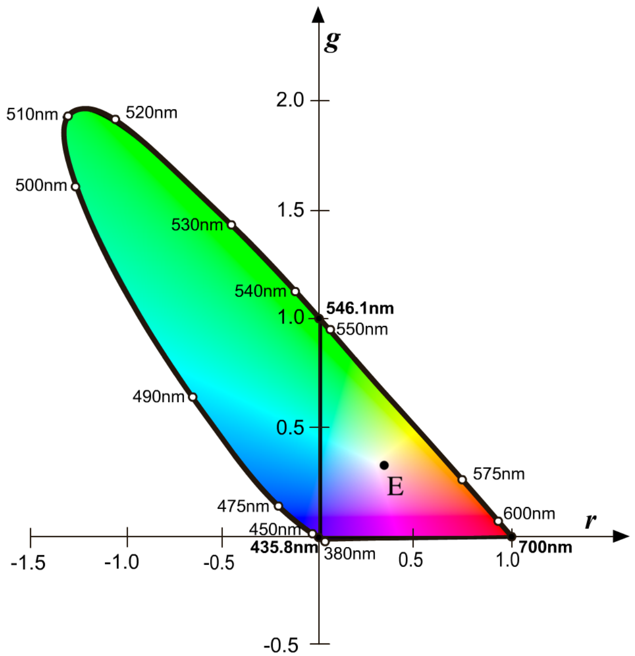


Fig. A.5 *C.I.E. (r, g)* chromaticity diagram corresponding to the *1931 C.I.E. standard observer* shown in figure A.4

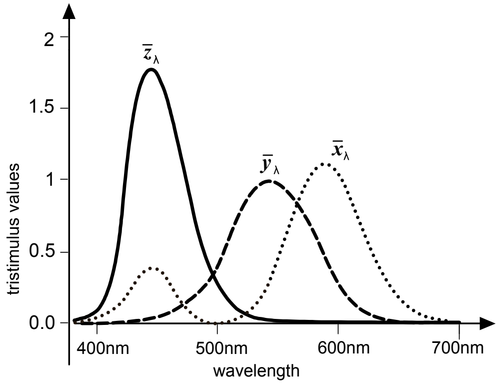


Fig. A.6 Colour matching functions based on imaginary primaries defining the *1931 C.I.E. standard observer*; source: [5]

Another issue was that it would be practical to have one of the matching functions represent *the photopic luminous efficiency function V(λ)* for the so called *CIE standard photopic observer* (CIE 1926). Photopic luminous efficiency is also referred to as luminosity function or sensitivity function. It specifies the ratio of the radiant power of a monochromatic stimulus of wavelength λ_m to that of a monochromatic stimulus of wavelength λ, when the two monochromatic stimuli produce equivalent luminous sensations under certain observation conditions. The function is given normalized based on the maximum value of the ratio.

A.1.4. C.I.E. (x.y) chromaticity diagram

When one of the matching functions represents the luminous efficiency function of the *CIE standard photopic observer*, the value of the corresponding trichromate directly provides the *luminance* of a color. This way the other two numbers can be used to specify the *chromaticity* of the color.

Based on these considerations the C.I.E. took care that the function originally belonging to green was modified in such a way that it represents the luminance. So the C.I.E. replaced the original matching functions shown in figure A.4 by another set of matching functions shown in figure A.6. It is noted that the colour matching functions are dimensionless conversion factors from spectral power distributions to their tri-stimulus representation [6]

We note that all the matching functions in figure A.6 have positive values. Therefore using the functions positive tristimulus values are obtained from an SPD. The values have the labels *X, Y,* and *Z.* These values are obtained in the same way as for *R, G, B* earlier. Namely the SPD is multiplied by the weighting function and the integral for each resulting function with respect to the visible spectrum yields the tristimulus value specifying the colour. Explicitly this is given by

$$\begin{aligned} X &= \int_0^\infty \bar{x}_\lambda(\lambda)\ P(\lambda)\,d\lambda \\ Y &= \int_0^\infty \bar{y}_\lambda(\lambda)\ P(\lambda)\,d\lambda \\ Z &= \int_0^\infty \bar{z}_\lambda(\lambda)\ P(\lambda)\,d\lambda \end{aligned} \tag{A.4}$$

In Eq. (A.4) $P(\lambda)$ is the spectral density of radiant flux per unit wavelength interval specified through the SPD of the light; λ is visible wavelength. The dimensions of the radiant flux are those of power, and appropriate units are Watts.

Applying Grassman's *proportionality law* [8] in the same way as before for *R, G, B* we obtain chromaticity coordinates from the tristimulus values *X, Y, Z* given by

$$x = \frac{X}{X+Y+Z}$$
$$y = \frac{Y}{X+Y+Z} \qquad \text{(A.5)}$$
$$z = \frac{Z}{X+Y+Z}$$

The range of possible values for *X, Y,* and *Z* forms a colour space known as the *1931 C.I.E. XYZ* space. The chromaticity diagram of this space is space is shown in figure A.7. Note that in figure A.7 only the *x* and *y* coordinates are plotted.

As in the *C.I.E. RGB* space the colour located at $x_E=y_E=z_E=1/3$ belongs to the *equal-energy stimulus E*, which is white. The relationship between the chromaticity coordinates *r, g, b* and *x, y, z* of any colour stimulus are expressed by the following transformation Eq. (A.6) [5].

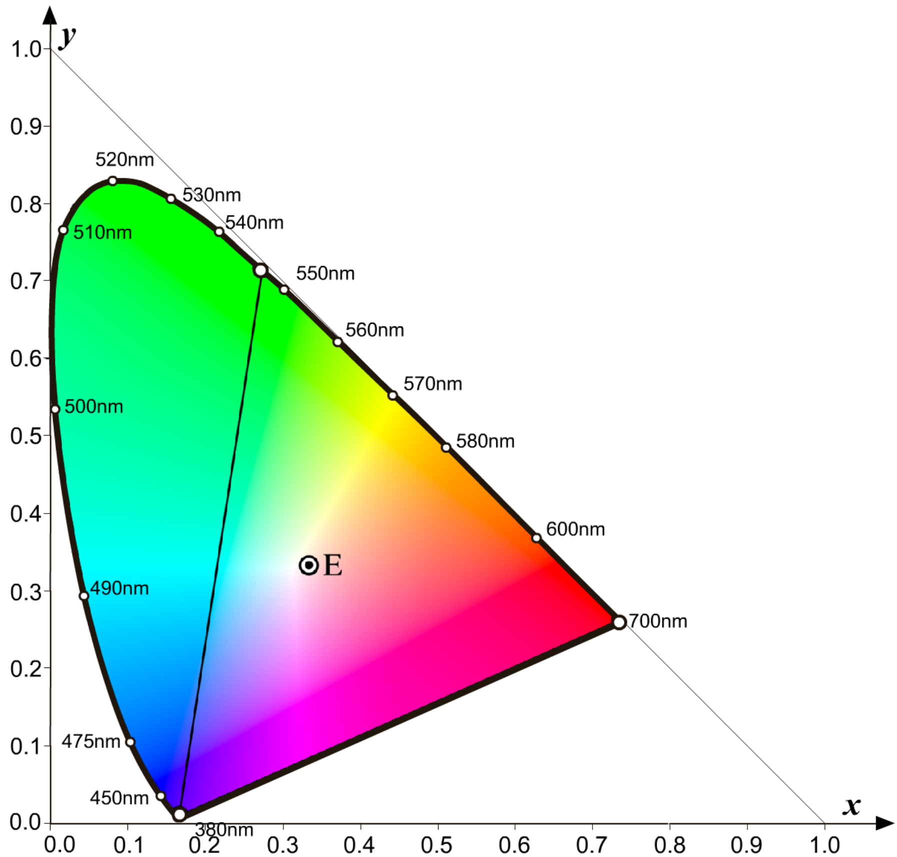


Fig. A.7 *C.I.E. (x, y)* chromaticity diagram corresponding to the *1931 C.I.E. XYZ* matching functions shown in figure A.6

$$x = \frac{0.49000r + 0.31000g + 0.20000b}{0.66697r + 1.13240g + 1.200663b}$$
$$y = \frac{0.17697r + 0.81240g + 0.01063b}{0.66697r + 1.13240g + 1.200663b} \quad \text{(A.6)}$$
$$z = \frac{0.00000r + 0.01000g + 0.99000b}{0.66697r + 1.13240g + 1.200663b}$$

Together the 1931 *C.I.E. RGB* and *XYZ* system form the *C.I.E. 1931 Standard Colorimetric System*. In the original experiments to obtain the *1931 C.I.E. standard observer* matching functions, the observers judged the colour matches having a field of view with a solid angle of 2°.

The human retina is non-homogenously populated with receptor cells. That is, the relative amount of the different cone types varies with eccentricity from the fovea. Therefore experiments with a field of view of 10° were conducted to supplement the earlier standard observer model [9].

Through the experiments new matching functions were obtained termed $\bar{r}_{10}(\lambda), \bar{g}_{10}(\lambda), and\, \bar{b}_{10}(\lambda)$, and in the same way as it was done with the *1931 C.I.E. standard observer* these functions are converted to *x, y, z* matching functions. The conversion is given by Eq. (A.7). It is noted that in the case for a 10° viewing angle the condition that the matching function for *y* represents the luminous efficiency function of the *CIE standard photopic observer* is not fulfilled.

$$\bar{x}_{10}(\lambda) = 0.341080\,\bar{r}_{10}(\lambda) + 0.189145\,\bar{g}_{10}(\lambda) + 0.387529\,\bar{b}_{10}(\lambda)$$
$$\bar{y}_{10}(\lambda) = 0.139058\,\bar{r}_{10}(\lambda) + 0.837460\,\bar{g}_{10}(\lambda) + 0.073316\,\bar{b}_{10}(\lambda) \quad \text{(A.7)}$$
$$\bar{z}_{10}(\lambda) = 0.000000\,\bar{r}_{10}(\lambda) + 0.039553\,\bar{g}_{10}(\lambda) + 2.026200\,\bar{b}_{10}(\lambda)$$

The new model was termed the *C.I.E. 1964 Supplementary Standard Colorimetric System*. Specification of the SPD of a colour by means of *C.I.E. 1964* tristimulus values or chromaticities is done in the same way as before, i.e. as given by Eq. (A.4).

A.2. Colour difference

Colour difference is the difference between two colours. Having specified two colours as two points in a three dimensional colour, space such as *CIE XYZ* or *CIE RGB space,* intuitively we may consider that the Euclidian distance measured between the two points indicates the difference between the colours. However for this to be the case it is necessary that the colour space obeys a certain criterion. This criterion is known as *perceptual uniformity*.

A.2.1. Perceptual uniformity

A colour space is termed perceptually uniform, when the colour difference a person with normal vision experiences between any two colours specified in this space is equal to the Euclidian distance measured between the colour coordinates.

Both standard colour spaces *CIE XYZ* and *CIE RGB* are not perceptually uniform. This has been shown by MacAdam [10]. MacAdam conducted colour matching experiments, where the stimuli had the same constant luminance of about 48cd/m^2 throughout the experiment. He investigated what difference in *x, y* colour chromaticity is just noticeable for humans. In the experiment observers modified the chromaticity of a variable stimulus along a straight

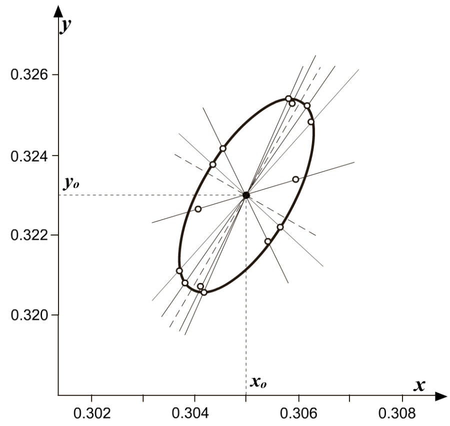


Fig. A.8 A Mac Adam Ellipse in CIE 1931 (x, y)-chromaticity diagram: the standard deviations of observers' colour matching judgements with respect to comparison chromaticity *(x_o, y_o)* [10]

line in the *1931 CIE x, y,* chromaticity diagram that passes through a certain *x, y* chromaticity coordinate. This coordinate specified the stimulus against which the variable stimulus was compared to find a match. MacAdam plotted the locations of the resulting standard deviations with respect to the comparison chromaticity. This is shown for the location $x_o=0.305$, $y_o=0.323$ in figure A.8. The standard deviations are located roughly on an ellipse shape that is centred at x_o, y_o. This is seen from figure A.8. Such an ellipse is termed *MacAdam ellipse*.

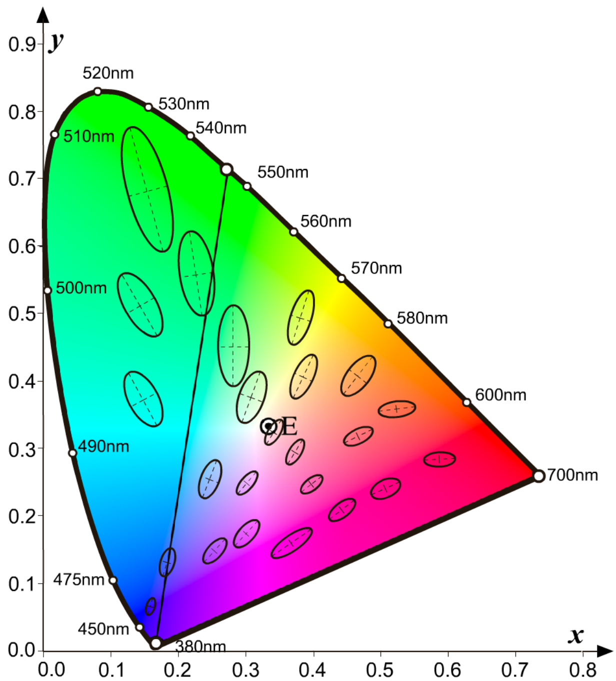


Fig. A.9 Mac Adam Ellipses indicate perceptual non-uniformity of the 1931 CIE x,y, chromaticity diagram [10]

Figure A.9 shows a number of MacAdam ellipses plotted in the *1932 CIE (x, y)* chromaticity diagram. For better visual appreciation of the differences, the sizes of the ellipses shown in the figure are enlarged by factor 10 with respect to the computed values of the standard deviations of the colour matching.

As can be seen from figure A.9, clearly colour difference appreciation in the 1932 XYZ space is not uniform. For the same distance among two colours in the chromaticity diagram the colours appear to be more similar in the green region versus the red region of the space. In a perceptually uniform colour space the standard deviations of distance estimations should not form ellipses but circles of the same size anywhere in the corresponding chromaticity diagram. To reconcile this issue the C.I.E. established two approximately perceptually uniform colour spaces, named *CIE 1976* $L^*u^*v^*$ and *CIE 1976* $L^*a^*b^*$-Space that are obtained from the CIE XYZ space.

*A.2.2. C.I.E. 1976 $L^*u^*v^*$ and L*a*b* Space*

The first approximately perceptually uniform colour space is named the *CIE 1976* $L^*u^*v^*$-Space. The second one established by the *C.I.E.* is named the *CIE 1976* $L^*a^*b^*$-Space [5].

Having specified two colors in one of *these spaces* the color difference ΔE^*_{uv} between the colors is computed as the Euclidian distance between their chromaticity coordinates. This is given by Eq. (A.8) for the $L^*u^*v^*$-space and by Eq. (A.9) for the $L^*a^*b^*$-Space.These equations are knows as *the color difference formulae.*

$$\Delta E^*_{uv} = \sqrt{(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2} \tag{A.8}$$

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \tag{A.9}$$

The *L** function defines the lightness correlate in both *CIE L*u*v** and *L*a*b** spaces. No simple relation exists between the chromaticity coordinates *u*, v** and *a*, b**.

We note that for constant *L** the chromaticity coordinates *u** and *v** space are a projective transformation of the *x, y* chromaticities in *CIE 1931 (x, y)* space. This means a straight line in a *CIE 1931 (x, y)* diagram remains a straight line in the *(u*, v*)* diagram.

A.3. Conversion from RGB to CIE L*a*b* space

A.3.1. Objective

During architectural design colours are often displayed on a computer monitor and they are coded in a RGB colour space. In order to compute the perceived colour difference between two colour samples given as RGB values it is necessary to express these colours in perceptually uniform spaces $L^*a^*b^*$ or $L^*u^*v^*$. For this purpose, first the RGB colours have to be transformed into *C.I.E. XYZ* space and thereafter to $L^*a^*b^*$ or $L^*u^*v^*$ space. In order to accomplish this first step correctly, certain properties of computer monitors need to be taken. Next to that the chromaticity characteristics of the RGB system and the white point representing the illumination condition should be taken into account. This is explained below.

A.3.2. Gamma correction

The luminance generated by a physical device is generally not a linear function of the applied signal strength. A conventional CRT monitor has a power-law response to voltage.

This means the luminance produced at the display surface is related to the applied voltage as given by Eq. (A.10) [11]

$$V_{out} = V_{in}^{\gamma} \tag{A.10}$$

where γ characterizes the non-linearity. A typical γ value is γ =2.2 for CRTs. To account for this images displayed on monitors are usually stored with *gamma correction.* This means the original linear RGB values of each pixel have been modified, so that they become $R_c=R^{1/2.2}$, $G_c=G^{1/2.2}$, $B_c=B^{1/2.2}$ respectively. Therefore, before being able to convert RGB values of an image displayed on a monitor to tristimulus values in XYZ space the non-linear RGB values should be converted to linear ones, so that the gamma correction is reversed. Explicitly this is given by Eq. (A.11)

$$\begin{aligned} r &= R_c^{\gamma} \\ g &= G_c^{y} \\ b &= B_c^{\gamma} \end{aligned} \tag{A.11}$$

where the R_c G_c B_c colour space is the unit cube, so that the RGB values range between zero and one.

A.3.3. Transform from linear RGB to XYZ space

Having reversed the gamma correction RGB values can be converted to XYZ tristimulus values using a matrix multiplication as given by Eq. (A.12)

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} r \\ g \\ b \end{bmatrix} [M] \tag{A.12}$$

The transformation matrix *M* is calculated from the RGB reference primaries of the RGB colour space. There is not one unique conversion matrix, because there are several common RGB spaces. Each one is based on its own RGB primaries. This means every *R G B* primary can be specified by a different chromaticity coordinate in the *C.I.E. (x, y)* chromaticity diagram. Next to that every *RGB* space has its own reference white specified by certain *XYZ* tristimulus values. Usually the white is selected as one of the CIE standard illuminants. Given the chromaticity coordinates of an RGB system (x_r, y_r), (x_g, y_g) and (x_b, y_b) and its reference white (X_W, Y_W, Z_W), the transformation matrix *M* is obtained as given by (A.14) and (A.15) [12].

Table A.1 Basic colours of the visible spectrum

	L^*	a^*	b^*
Red	53	66	52
Orange	68	37	73
Yellow	92	-6	90
Green	64	-44	48
Blue	42	28	-57
Violet	40	45	-38
White	96	-3	0
Black	9	2	3

$$[M] = \begin{bmatrix} S_r X_r & S_r Y_r & S_r Z_r \\ S_g X_g & S_g Y_g & S_g Z_g \\ S_b X_b & S_b Y_b & S_b Z_b \end{bmatrix}, \tag{A.13}$$

where

$$\begin{array}{lll} X_r = x_r / y_r & X_g = x_g / y_g & X_b = x_b / y_b \\ Y_r = 1 \quad , & Y_g = 1 \quad , & Y_b = 1 \\ Z_r = (1 - x_r - y_r) / y_r & Z_g = (1 - x_g - y_g) / y_g & Z_b = (1 - x_b - y_b) / y_b \end{array}$$

$$\begin{bmatrix} S_r \\ S_g \\ S_b \end{bmatrix} = \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \begin{bmatrix} X_r & Y_r & Z_r \\ X_g & Y_g & Z_g \\ X_b & Y_b & Z_b \end{bmatrix}^{-1} \tag{A.14}$$

A.3.4. Mean colour difference

Aiming to integrate the colour perception with the geometric component of visual perception, a certain value characterizing a colour space is significant. This is the mean colour difference. To obtain this mean value colour difference computations are conducted for the basic colours of the colour spectrum. Basic colours and their specification in *CIE* $L^*a^*b^*$ coordinates is given in table A.1.

The results of the colour difference computation among these basic colours are given in table A.2. The mean of the differences between a certain basic colour and the remaining basic colours are presented in table A.3. We note that yellow and blue are the colours with the highest mean difference with the other basic colours. White, orange and black have the least difference. A characteristic value for colour perception is the sum of all combinations of colour differences among the basic colours given in table A.4. This value is given by Eq. (A.15).

The mean color difference value is obtained by

Table A.2 Results of perceptually uniform colour difference computation for the basic colours given in table A.2.

Color difference	ΔE_{ab}^*	Color difference	ΔE_{ab}^*
Red-Orange	39	Yellow-Blue	159
Red-Yellow	94	Yellow-Violet	147
Red-Green	111	Yellow-White	91
Red-Blue	116	Yellow-Black	123
Red-Violet	93	Green-Blue	129
Red-White	97	Green-Violet	126
Red-Black	92	Green-White	71
Orange-Yellow	56	Green-Black	85
Orange-Green	85	Blue-Violet	26
Orange-Blue	133	Blue-White	90
Orange-Violet	115	Blue-Black	74
Orange-White	88	Violet-White	83
Orange-Black	98	Violet-Black	67
Yellow-Green	63	Black-White	87

$$\sum_{n=1}^{n=28} \Delta E_n^* \approx 2638 \quad \text{, so that}$$
$$\Delta \overline{E}_n^* \approx 2638 / 28 = 94 \tag{A.15}$$

This value plays an important role in human perception as this is explained below.

Table A.3 lour difference between a colour and the other basic colours of the spectrum

	Color	$\Delta \overline{E}^*$
1	Red	92
2	Orange	88
3	Yellow	105
4	Green	96
5	Blue	104
6	Violet	94
7	White	87
8	Black	89

A.4. Comparison of colour difference in chromatic versus greyscale scenes

A.4.1. Example images

As an illustrative example to investigate the difference of a coloured scene versus its greyscale counterpart a composition of colour patches reminding of the paintings by the Dutch artist Mondrian is chosen. The colour image and its greyscale counterpart are shown in figure A.10 and A.11 respectively. The colours in the image shown in figure A.10 are selected from the palette of basic colours given in table A.1.

The greyscale image in figure A.11 is obtained from the colour image by setting both chromaticity coordinates *a* and *b* of the $L^*a^*b^*$ colour specification to zero and keeping only

Fig. A.10 Colour image subject to computation of colour perception

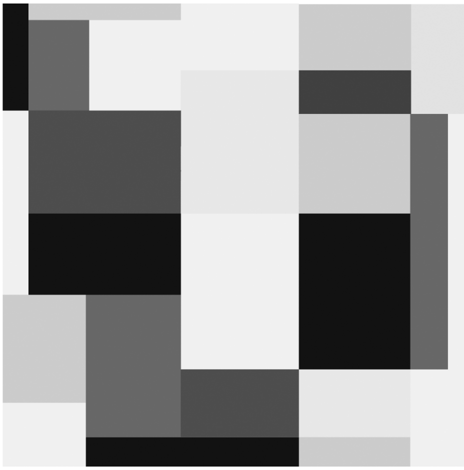


Fig. A.11 Greyscale version of the image shown in figure A.10 subject to computation of colour perception

the lightness component *L*. That is, in a greyscale case colour difference is equivalent to lightness difference.

A.4.2. Colour difference for the chromatic case

The colour perception at each border segment of the patches is computed by

$$P_c = \sum_{i=1}^{i=n} w_i \frac{\Delta E_i^*}{\overline{\Delta E^*}} = \frac{1}{\overline{\Delta E^*}} \sum w_i \, \Delta E_i^* \quad . \tag{A.16}$$

$\overline{\Delta E^*}$ is the mean colour difference with respect to all edges delimiting the objects forming the scene in question weighted by the factor w_i. It is noted that for general comparison purposes the value given by Eq. (A.15) may be used as mean colour difference in Eq. (A.16). The value *n* gives the total amount of edges delimiting the object. The factor w_i in Eq. (A.16) is given by

$$w_i = \frac{b_i}{b_1 + b_2 + ... + b_n} \tag{A.17}$$

where b_i is the length of the edge *i* in question. This means w_i gives the proportional length of a borderline segment b_i with respect to the sum of all borderlines of the scene.

The computation is shown in table A.4, where the border segments b_1-b_i are numbered starting with the top-left most edge of the respective patch and then following the patch outline clockwise. The length of each border segment is given in pixels in the table.

Table A.4 Colour perception for the patches in the chromatic image shown in figure A.10

		b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
p1	*ΔE**	91	159	96	91	161	91		
	length	95	34	77	95	81	31		
	w	.018	.006	.014	.018	.015	.006		
	P_c	.018	.012	.016	.018	.028	.006		
p2	*ΔE**	91	87	90	97	87			
	length	95	123	95	59	64			

	w	.018	.023	.018	.011	.012			
	P_c	.018	.023	.018	.012	.012			
p3	*ΔE**	80	93	77	96				
	length	90	79	90	79				
	w	.017	.015	.017	.015				
	P_c	.015	.015	.015	.016				
p4	*ΔE**	118	90	161	74	90			
	length	49	73	81	122	81			
	W	.009	.013	.015	.023	.015			
	P_c	.012	.014	.028	.019	.015			
p5	*ΔE**	74	87	92	77	87			
	length	122	64	77	46	64			
	w	.023	.012	.014	.009	.012			
	P_c	.019	.012	.015	.008	.012			
p6	*ΔE**	80	161	80	161				
	length	90	34	90	34				
	w	.017	.006	.017	.006				
	P_c	.015	.012	.015	.012				
p7	*ΔE**	77	92	121	87				
	length	90	123	90	123				
	w	.017	.023	.017	.023				
	P_c	.015	.024	.023	.023				
p8	*ΔE**	92	97	118	92	97	93		
	length	77	59	53	77	27	85		
	w	.014	.011	.010	.014	.005	.016		
	P_c	.015	.012	.013	.015	.006	.017		
p9	*ΔE**	90	161	74	118				
	length	95	54	95	54				
	w	.018	.010	.018	.010				
	P_c	.018	.018	.015	.013				
p10	*ΔE**	121	91	96	161				
	length	90	54	90	54				
	w	.017	.010	.017	.010				
	P_c	.023	.010	.018	.018				
p11	*ΔE**	11	11	5	11	91	91	90	97
	length	74	14	95	52	95	32	74	71
	w	.014	.003	.018	.010	.018	.006	.014	.013
	P_c	.002	.000	.001	.001	.018	.006	.014	.015
p12	*ΔE**	93	97	118	92				
	length	49	71	49	71				
	w	.009	.013	.009	.013				
	P_c	.010	.015	.012	.014				
p13	*ΔE**	13	96	80	11				
	length	90	52	90	52				
	w	.017	.010	.017	.010				
	P_c	.002	.011	.015	.001				
p14	*ΔE**	95	97	97	92	93			
	length	30	201	30	123	79			
	w	.006	.037	.006	.023	.015			
	P_c	.006	.041	.006	.024	.015			
p15	*ΔE**	11	77	93	11	11			
	length	21	46	85	67	85			
	w	.004	.009	.016	.012	.016			
	P_c	.000	.008	.017	.002	.002			

A.4.3. Colour difference for the greyscale case

In the same way the colour perception computations are carried out for the greyscale image. The results of the computations are presented in table A.5.

Table A.5 Colour perception of the patches in the greyscale image shown in figure A.11

		b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
p1	*ΔE**	2	58	8	2	52	2		
	length	95	34	77	95	81	31		
	w	.018	.006	.014	.018	.015	.006		
	P_c	.001	.009	.003	.001	.019	.000		
p2	*ΔE**	2	87	54	43	87			
	length	95	123	95	59	64			
	w	.018	.023	.018	.011	.012			
	P_c	.001	.049	.024	.012	.026			
p3	*ΔE**	49	33	77	8				
	length	90	79	90	79				
	w	.017	.015	.017	.015				
	P_c	.020	.012	.032	.003				
p4	*ΔE**	11	54	52	33	54			
	length	49	73	81	122	81			
	w	.009	.013	.015	.023	.015			
	P_c	.002	.018	.019	.018	.020			
p5	*ΔE**	33	87	44	77	87			
	length	122	64	77	46	64			
	w	.023	.012	.014	.009	.012			
	P_c	.018	.026	.015	.016	.026			
p6	*ΔE**	49	55	49	58				
	length	90	34	90	34				
	w	.017	.006	.017	.006				
	P_c	.020	.009	.020	.009				
p7	*ΔE**	77	44	85	87				
	length	90	123	90	123				
	w	.017	.023	.017	.023				
	P_c	.032	.025	.035	.049				
p8	*ΔE**	44	43	11	44	43	33		
	length	77	59	53	77	27	85		
	w	.014	.011	.010	.014	.005	.016		
	P_c	.015	.012	.003	.015	.005	.013		
p9	*ΔE**	54	62	33	11				
	length	95	54	95	54				
	w	.018	.010	.018	.010				
	P_c	.024	.015	.014	.003				
p10	*ΔE**	85	2	8	62				
	length	90	54	90	54				
	w	.017	.010	.017	.010				
	P_c	.035	.000	.003	.015				
p11	*ΔE**	10	10	4	10	2	2	54	43
	length	74	14	95	52	95	32	74	71
	w	.014	.003	.018	.010	.018	.006	.014	.013
	P_c	.003	.001	.002	.002	.001	.000	.018	.014
p12	*ΔE**	33	43	11	43				
	length	49	71	49	71				
	w	.009	.013	.009	.013				

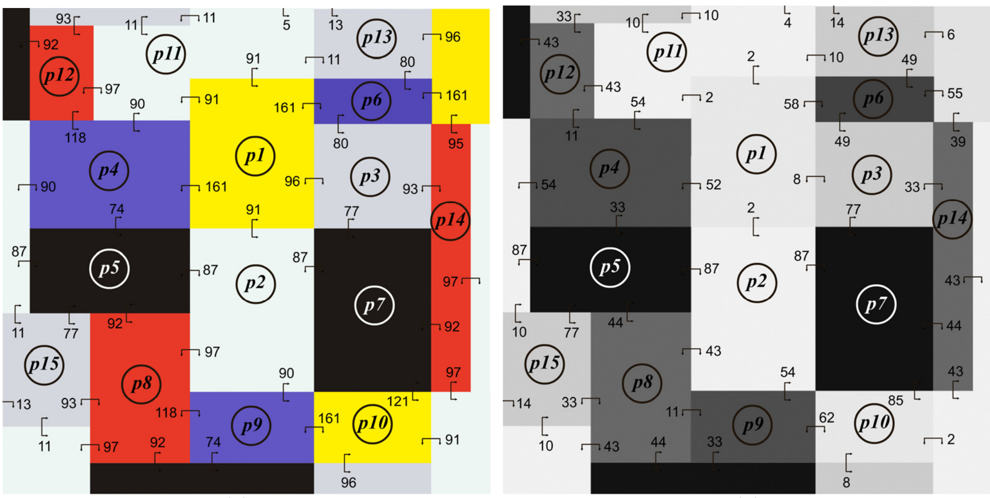
	P_c	.007	.014	.002	.014	
p13	*ΔE**	14	6	49	10	
	length	90	52	90	52	
	w	.017	.010	.017	.010	
	P_c	.006	.001	.020	.002	
p14	*ΔE**	39	43	43	44	33
	length	30	201	30	123	79
	w	.006	.037	.006	.023	.015
	P_c	.005	.040	.006	.025	.012
p15	*ΔE**	10	77	33	10	14
	length	21	46	85	67	85
	w	.004	.009	.016	.012	.016
	P_c	.001	.016	.013	.003	.005

A.4.4. Results

The resulting colour differences ΔE_{ab}^* among the patches in the image are shown in figures A.12a and A.12b for the colour and greyscale image respectively. Comparing the figures we note that the colour differences in the chromatic image are generally higher than in the greyscale image. The value of $\overline{\Delta E^*}$ is found to be 87.8 in the chromatic image. In the greyscale case $\overline{\Delta E^*}$ is found to be 40.3.

A.4.5. Chromaticity factor

The ratio between the mean colour differences for each scene gives a factor indicating the relative difference between the chromatic image and its greyscale version. This factor we term chromaticity factor and denote it by *C*. However, simply averaging the colour differences of the scene is a rough approach as it does not take into account the specific composition of the picture, i.e. its particular segmentation. Accurate computation of the factor indicating the significance of chromaticity with respect to greyscale is given by



(a) (b)

Fig. A.12 Colour differences between patches at their boundaries for the chromatic image (a); for the greyscale image (b)

$$C = \frac{\sum w_i \, \Delta E_i^*}{\sum w_i \, \Delta E_j^*} \tag{A.18}$$

where ΔE_i^* belongs to the chromatic image and ΔE_j^* belongs to the greyscale image.

In order to obtain the value of the chromaticity factor *C* for each patch *i* the *w ΔE** values of all borders of the patch are summed up and the ratio between chromatic and greyscale case is calculated as given by Eq. (A.18). The result is given in table A.6 for each patch respectively. The chromaticity factor for each patch is shown in the rightmost column of Table A.6. From figure A.13 it is noted that for patches 11 and 1 the difference of the colour perception between the greyscale and chromatic image is the largest. Namely it is about factor 13.3 and 6.3 respectively. This means that in the greyscale version of patch 11 is about 13 times less noticeable compared to the chromatic one. The greyscale version of patch 1 is about 6.3 times less noticeable compared to the chromatic version. For the other patches the factor ranges between 1.3 and 2.9. For patch 5 and 7 clearly there is the least perceptual difference comparing the greyscale and the colour image. This is because these patches have a black colour. Therefore their colour specification in *L*a*b** space is almost identical with their specification on the Lightness scale.The total chromaticity value *C* of the whole scene is obtained as

$$C = \frac{\sum w_i \, \Delta E_i^*}{\sum w_i \, \Delta E_j^*} \cong \frac{87.82}{39.01} \cong 2.25 \cdot$$

This means, concerning the whole image, chromaticity boosts perception more than two times compared to the greyscale case. This entails that in the present case the average colour perception P_c viewing a coloured image is more than twice the magnitude compared to viewing the greyscale version of the same image.

Table A.6 Chromaticity factors per patch

patch	Chromatic image $w_i \, \Delta E_i^*$	Greyscale image $w_i \, \Delta E_j^*$	C_i
1	8,53	1,35	6,34
2	7,25	4,46	1,62
3	5,36	2,69	1,99
4	7,72	3,17	2,44
5	5,69	4,09	1,39
6	4,72	2,36	2,00
7	7,36	5,67	1,30
8	6,78	2,56	2,65
9	5,65	2,26	2,51
10	6,11	2,18	2,80
11	5,02	0,38	13,33
12	4,41	1,53	2,87
13	2,59	1,21	2,14
14	8,14	3,54	2,30
15	2,48	1,56	1,59

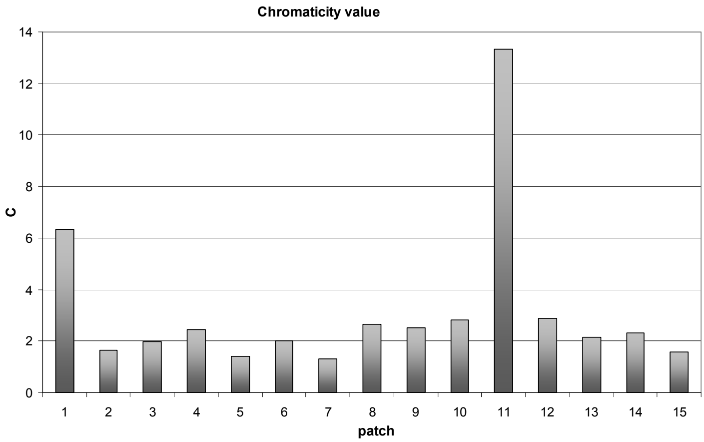


Fig. A.13 *C* values for each patch

A.4.6. Composite perception

In order to visually verify the perceptual difference between the colour and greyscale image let us compute the combined perception consisting of the colour perception and the geometry based perception of the patches.

For this purpose the geometric perception is obtained as follows: When the image is observed from a larger distance the probability density function describing the visual perception per unit length can be approximated by a uniform pdf. In the present case the image has a width and height of 1000 pixels, so that the $P_x(x)$ and $P_y(y)$ are both equal to 1/1000.

Table A.7 Resulting perception of the patches

		color image		**grayscale image**	
patch	P_g	P_c	P_C	P_c	P_C
p1	0.101	0,097	0,008	0,033	0,003
p2	0.111	0,083	0,007	0,111	0,012
p3	0.068	0,061	0,003	0,067	0,005
p4	0.095	0,088	0,006	0,079	0,007
p5	0.075	0,065	0,004	0,101	0,008
p6	0.030	0,054	0,001	0,058	0,002
p7	0.105	0,084	0,007	0,141	0,015
p8	0.082	0,077	0,005	0,064	0,005
p9	0.048	0,064	0,002	0,056	0,003
p10	0.046	0,070	0,002	0,054	0,002
p11	0.049	0,057	0,002	0,042	0,002
p12	0.033	0,050	0,001	0,038	0,001
p13	0.044	0,029	0,001	0,030	0,001
p14	0.059	0,093	0,004	0,088	0,005
p15	0.055	0,028	0,001	0,039	0,002
Σ	1.000	1.000		1,000	

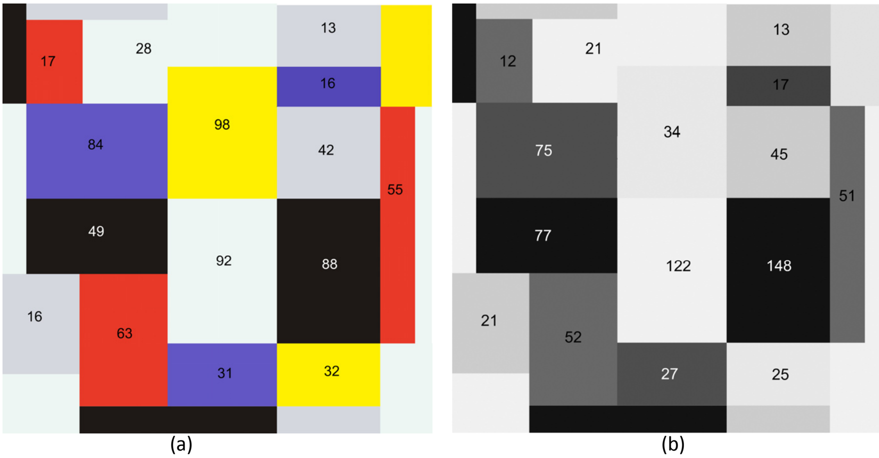


Fig. A.14 Degrees of composite perception P_C of the patches in the chromatic image (a) and its greyscale counterpart (b) scaled by factor 10^5

Therefore, the geometric perception component is computed by dividing the surface area of a patch by the total size of the image, i.e. by 10^6 square pixels. This is given by

$$P_C = P_g\, P_c = P_g \frac{1}{\Delta E^*} \sum w_i\, \Delta E_i^* , \qquad \text{(A.19)}$$

Where the resulting geometric perception P_g is then multiplied with the colour perception component P_c (see also Chapter 2). The resulting composite perception P_C is given in table A.7. The resulting perception values given in table A.7 are also given in figure A.14 at the respective patch they belong to. This is done so that the results can be verified easily. The numbers given in figure A.14a and A.14b are the perception values scaled by factor 10^5 for legibility. From the figure we note that In the chromatic case the large yellow patch (number 1) has the highest perception and patches 2, 4, and 7 are also high. In the greyscale case the large black patch (number 7) has the highest perception, and the large white patch (number 2) is also high.

A.5. Details on the interior perception analysis in Chapter 2

Table A.8 Circumferences of the objects in the scene.

Object	width	height	circumference
O1	450	700	1806
O2	250	220	738
O3	480	100	911
O4	130	185	495
O5	110	130	377
$O6_a$	250	115	573
$O6_b$	85	95	283
O7	210	230	691
O8	480	220	1099

Table A.9 Computation of colour perception P_c at each object border

		b1	b2	b3	b4
O1	*ΔE**	5.00E+01	6.00E+01		
	length	1.26E+03	5.42E+02		
	w	1.81E-01	7.77E-02		
	P_c	0.17	0.09		
O2	*ΔE**	9.70E+01	8.20E+01	8.80E+01	
	length	2.21E+02	3.69E+02	1.48E+02	
	w	3.18E-02	5.29E-02	2.12E-02	
	P_c	0.06	0.08	0.03	
O3	*ΔE**	3.80E+01	7.10E+01	8.80E+01	3.70E+01
	length	9.11E+01	2.28E+02	1.82E+02	4.10E+02
	w	1.31E-02	3.27E-02	2.61E-02	5.88E-02
	P_c	0.01	0.04	0.04	0.04
O4	*ΔE**	3.00E+01	3.80E+01	2.10E+01	
	length	2.97E+02	9.89E+01	9.89E+01	
	w	4.26E-02	1.42E-02	1.42E-02	
	P_c	0.02	0.01	0.01	
O5	*ΔE**	6.50E+01	4.00E+01		
	length	2.64E+02	1.13E+02		
	w	3.78E-02	1.62E-02		
	P_c	0.04	0.01		
O6a	*ΔE**	1.50E+01	2.70E+01	5.00E+01	2.60E+01
	length	1.15E+02	1.72E+02	8.60E+01	2.01E+02
	w	1.64E-02	2.47E-02	1.23E-02	2.88E-02
	P_c	0.00	0.01	0.01	0.01
O6b	*ΔE**	6.03E+01	8.50E+01	3.00E+01	
	length	5.65E+01	5.65E+01	1.70E+02	
	w	8.11E-03	8.11E-03	2.43E-02	
	P_c	0.01	0.01	0.01	
O7	*ΔE**	5.30E+01	5.50E+01	7.10E+01	
	length	3.45E+02	2.42E+02	1.04E+02	
	w	4.95E-02	3.47E-02	1.49E-02	
	P_c	0.05	0.03	0.02	
O8	*ΔE**	5.60E+01	6.20E+01		
	length	4.95E+02	6.04E+02		
	w	7.09E-02	8.67E-02		
	P_c	0.07	0.10		

References

1. Williamson, S.J., Cummins, S.Z.: Light and Color in Nature and Art. Wiley, New York (1983)
2. Stockman, A., MacLeod, D.I.A., Johnson, N.E.: Spectral sensitivities of the human cones. Opt. Soc. Am. 10 (1993) 2491-2521
3. Wright, W.D.: A re-determination of the trichromatic coefficients of the spectral colours. Transactions of the Optical Society 30 (1928) 141-164
4. Guild, J.: The colorimetric properties of the spectrum. Philosophical Transactions of the Royal Society of London A230 (1931) 149-187
5. Wyszecki, G., Stiles, W.S.: Color Science. Wiley, New York (1982)

6. Brill, M.H.: Do tristimulus values have units? Color Research & Applications 21 (1995) 310-313
7. CIE: Colorimetry - Publication CIE 15.2-1986. Central Bureau of the Commission Internationale d'Éclairage, Vienna (1986)
8. Grassmann, H.: Zur Theorie der Farbenmischung. Poggendorf', Ann. Phys. 89 (1853)
9. Stiles, W.S., Burch, J.M.: Color matching investigations: final report (1958). Optica Acta 6 (1959)
10. MacAdam: Visual sensitivities to color differences in daylight. J. Opt. Soc. Am. 32 (1942)
11. Poynton, C.: A Technical Introduction to Digital Video. Wiley, New York (1996)
12. Lindbloom, B.: http://brucelindbloom.com/. (2007)

List of publications

Book chapters

Bittermann, M.S., Sariyildiz, I.S., Ciftcioglu, Ö.: Blur in Human Vision and Increased Visual Realism in Virtual Environments. Lecture Notes on Computer Science. Springer Verlag (2007)

Ciftcioglu, Ö., Bittermann, M.S.: Adaptive formation of Pareto front in evolutionary multi-objective optimization. In: Lazinica, A. (ed.): Evolutionary Computation. In-Tech, Vienna (2009)

Ciftcioglu, O., Bittermann, M.S.: From perceptual towards cognitive robotics in the framework of evolutionary computation. In: Pennacchio, S. (ed.): Emerging Technologies, Robotics and Control Systems. InternationalSAR, Palermo, Italy (2009)

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Visual perception theory underlying perceptual navigation. Emerging Technologies, Robotics and Control Systems International Society for Advanced Research (2007) 139-153

Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, O.: Perception & Architecture. In: Bekkering, H., Hauptmann, D., Heijer, A.d., Klatte, J., Knaack, U., Manen, S.v. (eds.): The Architecture Annual 2005-2006 Delft University of Technology. 010 Publishers, Rotterdam (2007) 104-109

Journal papers

Bittermann, M.S., Ciftcioglu, Ö.: Visual perception model for architectural design. Journal of Design Research 7 (2008) 35-60

Bittermann, M.S., Sariyildiz, I.S., Ciftcioglu, Ö.: Visual perception in design and robotics. Integrated Computer-Aided Engineering 14 (2007) 73-91

Ciftcioglu, O., Bittermann, M.S.: From perceptual towards cognitive robotics in the framework of evolutionary computation. Int. J Fatory Automation, Robotics and Soft Computing (2009)

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Visual perception theory underlying perceptual navigation. Int. J. Factory Autom., Robotics and Soft Comp. (2007) 171-185

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Multiresolutional fusion of perceptions applied to robot navigation. Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII) 11 (2007) 688-700

Conference papers

Bittermann, M.S., Ciftcioglu, O.: A cognitive system based on fuzzy information processing and multi-objective evolutionary algorithm. IEEE Conference on Evolutionary Computation - CEC 2009. IEEE, Trondheim, Norway (2009)

Bittermann, M.S., Ciftcioglu, Ö.: Validation of a visual perception model. Joint Int. Conf. on Construction Culture, Innovation, and Management (CCIM), Dubai, United Arabian Emirates (2006) 289-299

Bittermann, M.S., Sariyildiz, I.S., Ciftcioglu, Ö.: Visual space perception model identification by evolutionary search. Design 2006, Dubrovnik, Croatia (2006) 185-192.

Bittermann, M.S., Ciftcioglu, O.: Real-time measurement of perceptual qualities in conceptual design. Proc. Int. Symp. Tools and Methods of Competitive Engineering TMCE 2006, Ljubljana, Slovenia (2006) 231-239

Bittermann, M.S., Ciftcioglu, O.: Systematic measurement of perceptual design qualities. Proc. Europ. Conf. Complex Systems ECCS 2005 workshop Embracing Complexity in Design, Paris, France (2005) 15-22

Bittermann, M.S.: Architectural design enhancement: Towards knowledge-based tools for intelligent, collaborative conceptual design. Int. Conf. Innovation in Architecture, Engineering and Construction, Rotterdam, The Netherlands (2005) 523-531

Ciftcioglu, O., M.S.Bittermann: Multiobjective Optimization for Cognitive Design. Joint 4th Int. Conf. on Soft Computing and Intelligent Systems (SCIS & ISIS), Nagoya, Japan (2008)

Ciftcioglu, Ö., Bittermann, M.S.: Solution Diversity in Multi-Objective Optimization: A study in Virtual Reality. World Congress on Computational Intelligence WCCI 2008, Hong Kong (2008)

Ciftcioglu, O., Bittermann, M.S., Sariyildiz, I.S.: Building performance analysis supported by GA. In: Tan, K.C., Xu, J.X. (eds.): 2007 IEEE Congress on Evolutionary Computation. IEEE, Singapore (2007) 489-495

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Sensor data fusion in autonomous robotics. The 2nd Int. Conf. Innov. Comp., Inf. and Contr. - ICICIC 2007. IEEE, Kumamoto, Japan (2007)

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Fuzzy neural tree for knowledge driven design. The 2nd Int. Conf. Innov. Comp., Inf. and Contr. - ICICIC 2007. IEEE, Kumamoto, Japan (2007)

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Further studies on visual perception for perceptual robotics. Fourth Int. Conf. Informatics in Control, Automation and Robotics - ICINCO2007, Angers, France (2007) 468-744

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: A neural fuzzy system for soft computing. NAFIPS 2007. IEEE, San Diego, USA (2007) 489-495

Ciftcioglu, O., Bittermann, M.S., Sariyildiz, I.S.: Fusion of perceptions for perceptual robotics. Proc. NAFIPS'06, Montréal, Québec, Canada. (2006)

Ciftcioglu, O., Bittermann, M.S., Sariyildiz, I.S.: Application of a visual perception model in virtual reality (poster). ACM SIGGRAPH Symp. Applied Perception in Graphics and Visualization, APGV'2006, Boston, USA (2006) 143

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Towards computer-based perception by modeling visual perception: a probabilistic theory. 2006 IEEE Int. Conf. on Systems, Man, and Cybernetics, Taipei, Taiwan (2006) 5152-5159

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Studies on visual perception for perceptual robotics. ICINCO 2006 - 3rd Int. Conf. on Informatics in Control, Automation and Robotics, Setubal, Portugal (2006)

Ciftcioglu, Ö., Bittermann, M.S., Sariyildiz, I.S.: Autonomous robotics by perception. SCIS & ISIS 2006, Joint 3rd Int. Conf. on Soft Computing and Intelligent Systems and 7th Int. Symp. on advanced Intelligent Systems, Tokyo, Japan (2006) 1963-1970

Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, Ö.: Multi-objective optimization in the construction industry. AEC 2008, Antalya, Turkey (2008)

Sariyildiz, I.S., Bittermann, M.S., Ciftcioglu, Ö.: Performance-based Pareto optimal design. In: Rusák, I.H.a.Z. (ed.): TMCE 2008, Izmir, Turkey (2008)

About the author

Michael S. Bittermann is an architect. He was born on 21st of July 1976 in Würzburg, Germany, into a family of builders. His father is a building engineer, who lead a design and manufacturing company specialized in passive houses, which he inherited from the candidate's grandfather. Bittermann received his secondary education in 1982-1995 from Matthias Grünewald Gymnasium in Tauberbischofsheim, Germany. He graduated with an overall score of 1.2 on the German scale (corresponding to 9.6 on the Dutch scale) and received the Physics award of the institution.

He studied architecture from 1997-2003. He graduated cum laude as Master of Science in Architecture from Delft University of Technology, in Delft, The Netherlands. He is a member of The Royal Institute of Dutch Architects (BNA). As practising architect he designed a number of residential buildings that were built in Germany by his father's company. During his architecture studies he was employed as an intern at the architecture firms 1100 Architect in New York, USA and ONL in Rotterdam, The Netherlands, working on various building projects ranging from interior design to the design of larger buildings. During his Masters studies he was a student assistant at the Hyperbody group of Delft University of Technology, supporting students in the design of interactive architecture, and co-supervising two architectural design workshops at International University of Catalunya (UIC), in Barcelona, Spain.

Mr. Bittermann conducted his PhD at the chair of Design Informatics, at the Architecture Faculty of Delft University of Technology. During his PhD he published thirty papers in the areas of architectural design and computational intelligence, including five chapters in international peer-reviewed books, and five papers in international peer-reviewed journals.

Privately Bittermann enjoys spending time with his wife and daughter. He also composes and performs music, where singing is his particular passion, and creates sculptures in stone.