Matching Representation to Perception

Building Identification Using a Mobile Architectural Guide

Alexander Koutamanis
Delft University of Technology, The Netherlands
http://caad.bk.tudelft.nl | http://www.re-h.nl
a.koutamanis@tudelft.nl

Visual representations play an important role in mobile architectural guides, especially with respect to the identification of buildings, i.e. matching internal representations to external ones (the documentation in the guide) and the perceived scene. By restructuring this documentation into networks of significant architectural entities and features a guide is enriched with flexible, economical means for supporting building identification under the typically variable conditions architectural travelers have to endure.

Keywords*: Mobile computing; architectural guide; representation; perception; recognition.*

Introduction: architectural guides

As all architectural travelers can testify, a good architectural guidebook is invaluable both as a navigation aid and as a succinct architectural manual. The information it contains on a building and its contexts (physical, social, historical, typological, morphological) supports identification and appreciation of the building. Unfortunately, up-to-date, comprehensive architectural guidebooks tend to be scarce. The combination of relatively low annual sales (in comparison to general tourist guidebooks) and high production cost (partly due to the necessity to include many illustrations) make the periodicity of architectural guidebooks quite low. As a result, a guidebook may miss the latest developments that nevertheless hold particular attraction to its audience.

Periodicity and the consequent problems of up-to-datedness derive from the production processes of analogue publications in general, which are to a degree constrained by the information container. Related to that are portability constraints, which lead to selectivity problems: in order to keep a guidebook compact, one may be forced to omit periods and styles or economize on (especially visual) information. This adds to the frequent mismatch between information supply and demand: architectural guidebooks cannot cater for the individual interests and preferences of each reader or for his possible lack of familiarity with specific periods, styles, cultures or places.

Such problems run contrary to developments in the digitization of information (Koutamanis, 1998). Electronic information carriers (especially online) offer not only a much lower periodicity but also a tremendous capacity and multimedia possibilities beyond the reach of analogue publications. Moreover, they shift selectivity from supply to demand: users are equipped with powerful and flexible retrieval tools for applying their own choice of subjects, media and appearance to the available information. Despite enduring copyright issues, there is abundant digital information on important buildings: texts, drawings and photographs but also video and 3D models. Online resources tend to focus on visual documentation. Just a few of them have been intended as architectural guides, with a complete presentation of a building and its context in mind (e.g. www.galinsky.com: May 2007). Nevertheless, the plethora of navigation services on the Internet makes the addition of visiting information to databases of buildings quite straightforward (e.g. www.greatbuildings.com: May 2007).

From a technical viewpoint a major step towards the development of digital architectural guides is the making of such information available to mobile computing devices. This relates to research into ubiquitous computing and smart buildings, where interaction with information and navigation are being extensively explored, especially in confined environments like museums (Bruns et al., 2005, Jeng, 2005, Nomura and Kishimoto, 2005, Shen and Teng, 2005). Interest in guided navigation with information feedback has been higher in other fields, with emphasis mostly on general computational and cognitive issues (Abowd et al., 1997, Cheverst et al., 2000, Persson et al., 2002). In recent years there has also been some interest in architectural guides on mobile devices but mostly from mobile telephony providers or related service providers (cultural or tourist). It is already possible to make architectural walks with one's own mobile phone providing textual or aural instructions and commentary in several cities (www.talktomenl.nl: May 2007, www.walkthetalk.hk: May 2007).

In CAAD research there has been so far only one directly relevant research project, thankfully with extensive ambitions and many facets (Berridge and Brown, 2002, Berridge et al., 2002, Brown et al., 2006). This project investigates the utility of digital documentation and explores the potential of mobile information processing (Berridge et al., 2003) but also considers the wider usability of existing representations towards more effective and efficient fundamental solutions to information and modeling problems (Knight et al., 2006).

MAG

The mobile architectural guide (MAG) is similarly motivated by the questions of how to make existing information available for mobile processing and what information is necessary for mobile architectural applications. MAG is a cross-platform modular system that has been tested on a variety of devices under several operating systems (Koutamanis, 2007 (forthcoming)). MAG comprises three kinds of modules:

- *Navigation* modules, largely based on commercially available systems. MAG accepts any navigation system but prefers satellite and hybrid terrestrial-satellite differential navigation systems (GPS and DGPS) because they offer higher precision without requiring a dedicated infrastructure.
- *Architectural information* modules: online and offline architectural information systems ranging from purpose-made multimedia databases on the mobile device to external online collections, which are connected to the MAG either directly (e.g. as a hyperlink to a web page) or through a linking database on the mobile device. Architectural information modules connect to navigation through practically universal overlays of points of interest (POIs).
- *Learning* modules that support interaction with MAG by registering, remembering and re-using user preferences and search patterns.

The combination of these modules results into mobile systems that allow architectural travelers to select a number of buildings they want to visit, plan their route to these buildings, and correctly identify the buildings once they have reached their destination. Alternatively architectural travelers are informed about possible architectural sights in the vicinity of their current location. If they are interested in these sights, MAG adds the selected sights to the intermediate waypoints of the existing route.

Architectural documentation and building perception

Initially the different versions of MAG were typical multimedia information systems: they combined available information in a particular application environment (mobile information processing) and augmented the capabilities of this environment so as to meet their particular requirements. They served well the two initial goals: (a) re-use of computer-based drawings and models in an educational environment and (b) presentation of information in a comprehensive yet straightforward manner. MAG was a digital, improved version of conventional analogue architectural guides.

A change of direction came with a critical observation made by almost all MAG test users: the architectural visual documentation was generally useful for understanding a building but did not always suffice for its identification. They suggested that the projections and conventions that appear to serve us in designing and constructing a building may be removed from the everyday perception of the building. This provided an incentive to explore the relationship between architectural representation and building perception in the well-defined context of MAG, i.e. with respect to building recognition and orientation in space. Consequently, the research focus shifted towards the use of visual information (drawings of all kinds, photographs, video) for the identification of buildings. The object of identification could be an architectural sight or a local landmark that facilitated orientation and route verification.

At the most fundamental level identification concerns meaningful objects, i.e. architectural entities like a column or a window of a particular type or style. This level involves implicit knowledge, e.g. that a window contains glass panes (even if there is no indication of glass panes in the representation), strong cues that e.g. the building is classical because of the morphology of the window or about the function of the building (e.g. public building because of the scale of a door or stairs). Related to this level is the recognition of crucial features, especially of larger objects, such as the corners of a volume and edge junctions in general (Clowes, 1971, Huffman, 1971, Waltz, 1975). In some cases architectural entities can be considered as features of larger configurations, e.g. the capitals in a classical column. Identification is strongly influenced by such entities at the cost of connecting parts such as the shaft of a column, which may be incomplete, malformed or even absent.

The results of this level are often sufficient for the recognition and categorization of the whole, i.e. in the case of MAG for the correct identification of a building. This may occur directly, without further processing of entities and features, but in many cases we have to make larger configurations and their properties explicit, e.g. recognize partially occluded surfaces and volumes or group a number of columns together into a colonnade on the basis of relationships of axial and translational symmetry.

The starting point for the investigation of building identification in MAG became the strong relationship between visual perception and imagery (Kosslyn, 1994). The main hypothesis was that the way architectural visual documentation was presented in MAG could be based on the common mechanisms and principles that underlie the formation and manipulation of mental images on the one hand and the perception of visual scenes on the other. The purpose of this was to facilitate matching of the internal representations used by the architectural traveler to both external representations (documentation) and the scene perceived visually. To achieve this goal MAG should support a number of abilities that are central to visual identification (Kosslyn, 1994):

- *Indifference to location and distance*: we are capable of perceiving the same window at different angles and distances on a building or standing and fallen columns in the ruins of a classical temple (perceptual constancy)
- *Indifference to shape variation* of the whole object, its parts, spatial relations between parts and having or missing optional parts

- *Insensitivity to partial occlusion and degradation* in an image, including being equally capable of perceiving an architectural entity in a photograph and in a line drawing or when standing very close to the object of our observation
- *Identification of specific instances:* we can identify not only the type of an object but also each specific instance (e.g. a column is not only Ionic but also the third from the left when viewing a particular temple from a specific point). This also includes specific typological, morphological and spatial relations between entities or features, e.g. differences in the detailing of a particular astragal or changes in the distances between columns in a peristyle.
- *Correlation and distinction of objects and scenes:* we rarely perceive isolated objects, normally we have to deal with configurations of several objects, with multiple entities forming our focus at any given moment: e.g. viewing simultaneously a colonnade and a particular column or a facade and the fenestration at a part of the façade.

Such abilities and their importance in the perception of the real world and its representations (e.g. architectural drawings) put emphasis on the parts of a scene or entity and their spatial interrelationships, which form multiple representations and abstraction levels, both internally and externally. These representations can be either propositional, consisting of symbols indicating relations (predicates) and objects, or depictive, i.e. pictorial, conveying meaning through their resemblance to an object and consisting of parts that can be defined arbitrarily and flexibly. This agrees with the use of modular representations covering multiple abstraction levels in computational systems (Marr, 1982). It also relates to the tendency in CAAD (as well as building industrialization) to develop the traditionally depictive architectural representations in the direction of propositional systems by means of decomposition into identifiable parts (meaningful entities) and binding relations or constraints (e.g. parameterization).

In MAG these considerations led to more attention for the various conditions in use than for the

Figure 1
Insensitivity to partial occlusion and degradation; identification of specific instances; correlation and distinction of objects and scenes

different types of images and projections. Initial evaluations indicated that most users found line drawings and even wireframe models sufficient for recognition and identification purposes, even though color and texture (i.e. the added value of photographs and video) were among the critical factors for identification, especially under difficult conditions such as views obscured by vegetation or similarity with nearby buildings. The poor relationship between most architectural projections and perception was also identified: even computer-made 3D models and animations often choose for viewpoints and settings that provide design overview rather match the viewpoint of a visitor. This delayed rather than impeded recognition of a building. A more worrying observation concerned the selectivity and resulting incompleteness of drawing and models used for presentation or construction, which often miss elements and details that may change the appearance of a building.

Features, constraints and networks

Investigation of the various options in MAG was conducted by means of a monocular head-mounted display (HMD) that connected to the palmtop device and projected the images on the palmtop screen on a part of the visual field of the user. The HMD was considered to be ergonomically superior to most palmtop screens, especially in terms of usability becuase it supported the direct juxtaposition of stored images to the perceived scene. For safety reasons a monocular display was preferred over binocular, transparent HMDs that allow superimposition of image and scene (Azuma, 1997).

In the initial use tests of MAG (with or without the HMD) it became evident that users appreciated panoramic views at critical points in a route (or alternatively short videos where camera movement or zoom indicated the direction they should take), especially if annotated with pragmatic information such as street names and route directions. However, such images (being unstructured depictive representations) focus more on overviews that act as background to recognition tasks than on information directly involved in building recognition. As a result, they were less useful for the representation and recognition of the buildings to be visited.

A more promising direction was the use of 3D models that integrated different images and projections (Koutamanis, 2003). The most elementary form was composites of existing 2D drawings in 3D space (figure 2). These allowed for a rather clear presentation of a building in which users could identify relationships underlying design features, as well as for a reasonable support to identification. The latter relied heavily on software that permitted flexible viewing of the 3D composites (e.g. DWF viewers). Especially in complex buildings and in buildings that required a large number of drawings (including details), the legibility of the composite was greatly enhanced by the ability to change viewpoint and turn layers or parts of the composite on and off.

The main virtue of 3D composites lay in that they made direct and effect use of existing digital documents. With a minimum of processing MAG was enriched with integrated (or at least correlated) visual representations. Their main limitation was that

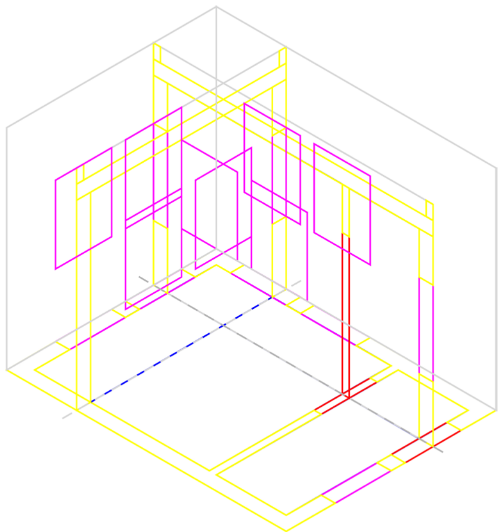

Figure 2
Integration of drawings and projections in 3D

they lacked an explicit framework, relying instead on perceived correspondences between different images. Abstract spatial representations of a building can provide such a framework, which moreover has strong relations with Euclidean and projective representations proposed in visual navigation studies, as well as with mental spatiotemporal representations of visual memories (Aloimonos et al., 1995, Fermüller and Aloimonos, 1995, Nelson, 1997, Riseman et al., 1997, Robert et al., 1997).

These abstract representations were implemented as simple spatial and volumetric X3D models which fitted the processing power and viewing capabilities of mobile devices. As with the 3D composites, the X3D models formed the basis for integrating the available visual information. This can be done in a direct manner, i.e. by mapping 2D images on the faces of a model (Stellingwerff, 2005). However, this has practical disadvantages (file complexity and size, image preprocessing) and does not add to the flexibility of the representation (in comparison with the viewers used for composites). An alternative that added abstraction capabilities was to concentrate on salient features and crucial entities, i.e. the objects and parts that determine recognition of an architectural scene.

By annotating X3D models with images of these features and entities the representation became a flexible and compact 3D network of meaningful parts. Each part could be an instance of a type (e.g. an edge junction or a window of a particular type). This caused a substantial reduction of redundancy, as the representation contained only one full description of each type. Instances of these types were anchored on explicit parts of the models (points, edges, faces). The resulting system was a multilevel representation comprising coordinating devices and elements. Coordinating devices could be global, i.e. abstract schemata and the overall form of a design, or local: constraints that focused on a particular element (Koutamanis, 1997).

The constraints of global and especially local coordinating devices can be implicit, i.e. inherent in the type of an element. For example, for each 2D or 3D corner type we can form connectivity expectations. These expectations identify not only the direction of connected corners but also their types (Koutamanis, 1995). The propagation network defined by such constraints is an efficient mechanism for correlation, verification and recognition of the entities the features belong to (Waltz, 1975). Explicit constraints generally refer to perceptual and composition

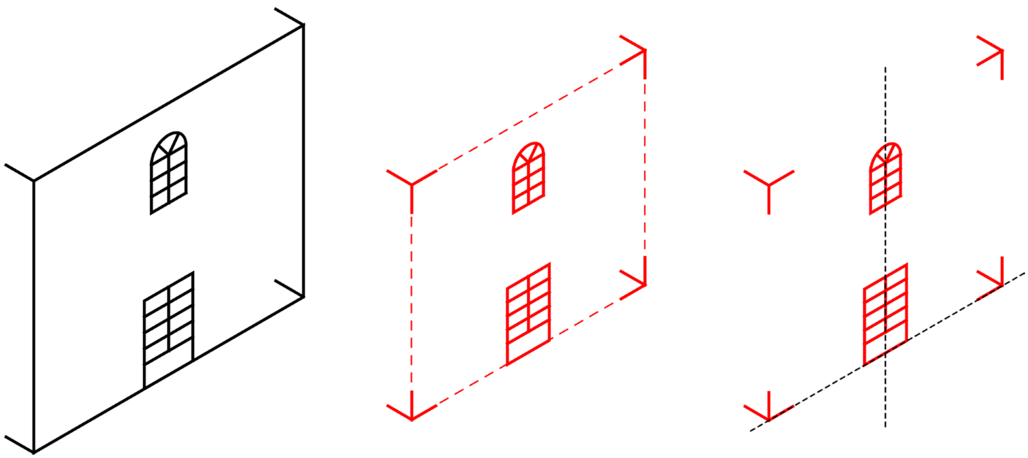

Figure 3
Features and relationships in an architectural object (left): features and constraints implicit in the features (middle); features and relationships of axial symmetry and alignment (left)

relations, primarily alignment and symmetry. Such relations can be part of global coordinating devices (e.g. the tripartition schema of classical architecture) or ad hoc conditions relating to design choices and construction or functional constraints (typical alignment sources).

These representations provide depictive and propositional information that not only facilitates recognition and identification of a building but also make explicit entities and relationships that support understanding of the building. In perceptual terms the main advantage of the representations is the transformability of the networks so as to account for different viewing conditions (also with respect to spatial resolution and conceptual specificity). The resulting images may be perspectively incorrect with respect to size and foreshortening but they aid recognition by providing information that is often suppressed in perspectively correct images (i.e. more detailed descriptions of salient features and elements than customary for a particular projection or scale) while allowing for direct comparison with the current view of the user. This was confirmed in a relative evaluation with raw and perspectively corrected photographs.

Discussion

In the technically restricted environment of MAG the performance of conventional analogue and digital architectural representations is significantly lower than for the original purposes of these representations. While not inappropriate for recognition and identification tasks, they provide little specific support for such tasks, especially under critical conditions. The decomposition of architectural representations into relations, constraints, features, architectural entities and the subsequent re-structuring of these into multilevel representations of coordinating devices and elements provides the required specificity and flexibility. These multilevel structures effectively segment depictive representations along propositional lines but on the basis of perceptual principles rather than the constructional/conceptual basis customary in digital design representations, from the original industrialization-motivated systems to recent interoperability standards. This perceptual segmentation also paves the way for the application of feature-based systems that can automatically identify artifacts by distinguishing between members of one or more classes (Föckler et al., 2005).

References

Abowd, G. D., Atkeson, C. G., Hong, J., Long, S., Kooper, R. and Pinkerton, M.: 1997, Cyberguide: A mobile context-aware tour guide, Wireless Networks, 3(5), pp. 421-433.

Aloimonos, Y., Fermüller, C. and Rosenfeld, A.: 1995, Seeing and understanding: representing the visual world, ACM Computing Surveys, 27(3), pp. 307-309.

Azuma, R. T.: 1997, A survey of augmented reality, Presence: Teleoperators and Virtual Environments, 6(4), pp. 355-385.

Berridge, P., Brown, A. and Knight, M.: 2002, One city to go, in A. M. E. Rafi, C. W. Khong, M. Neo, K. T. Neo and S. N. A. S. Ahmad (eds), CAADRIA 2002. Proceedings of the 7th Conference on CAAD Research in Asia, CAADRIA, Cyberjaya, pp. 57-64.

Berridge, P., Koch, V. and Brown, A. G. P.: 2003, Information spaces for mobile city access, International Journal of Architectural Computing, 1(1), pp. 34-45.

Berridge, P. and Brown, A.: 2002, A touring machine, Connecting the Real and the Virtual - design education. 20th eCAADe Conference Proceedings, eCAADe, Warsaw, pp. 488-493.

Brown, A., Knight, M., Chen, Y.-H. and Saeed, G.: 2006, City information delivered to mobile digital devices - reflection on contemporary potentail and problems, in V. Bourdakis and D. Charitos (eds), Communicating space(s). eCAADe, Volos, pp. 146-150.

Bruns, E., Brombach, B., Zeidler, T. and Bimber, O.: 2005 Enabling mobile phones to support large-scale museum guidance. Bauhaus University Weimar.

Cheverst, K., Davies, N., Mitchell, K., Friday, A. and Efstratiu, C.: 2000, Developing a context-aware electronic

tourist guide: some issues and experiences, Proceedings of the CHI'00. ACM international conference on human factors in computing systems, ACM Press, New York, pp. 17-24.
Clowes, M.: 1971, On seeing things, Artificial Intelligence, 2, pp. 79-116.
Fermüller, C. and Aloimonos, Y.: 1995, Vision and action, Image and Vision Computing, 13(10), pp. 725-744.
Föckler, P., Zeidler, T., Brombach, B., Bruns, E. and Bimber, O.: 2005, PhoneGuide: museum guidance supported by on-device object recognition on mobile phones, International Conference on Mobile and Ubiquitous computing (MUM'05). ACM Press, New York, pp. 3-10.
Huffman, D.: 1971, Impossible objects as nonsense sentences, in B. Meltzer and D. Michie (eds), Machine Intelligence, Edinburgh, Edinburgh University Press.
Jeng, T.: 2005, Advanced ubiquitous media for interactive space, in B. Martens and A. Brown (eds), CAAD Futures 2005, Kluwer, Dordrecht, pp. 341-350.
Knight, M. W., Brown, A. G. P. and Smith, J. S.: 2006, Digital terrain meshes from GPS in urban areas: A practical aid to city modelling, in A. Kaga and R. Naka (eds), CAADRIA 06. Proceedings of the 11th Conference on CAAD Research in Asia, CAADRIA, Kumamoto, pp. 443-451.
Kosslyn, S. M.: 1994, Image and brain: the resolution of the imagery debate, MIT Press, Cambridge, Massachusetts.
Koutamanis, A.: 1995, Recognition and retrieval in visual architectural databases, in A. Koutamanis, H. Timmermans and I. Vermeulen (eds), Visual databases in architecture. Recent advances in design and decision making, Avebury, Aldershot.
Koutamanis, A.: 1997, Multilevel representation of architectural designs, in R. Coyne, M. Ramscar, J. Lee and K. Zreik (eds), Design and the net, Europia Productions, Paris.
Koutamanis, A.: 1998, Information systems and the Internet: towards a news counter-revolution?, 4th Design and Decision Support Systems in Architecture and Urban Planning Conference, Eindhoven.
Koutamanis, A.: 2003, Integrale bouwkundige informatieverwerking en -beheer, in P. G. Luscuere (ed), Informatie-uitwisseling en -beheer bij grote bouwprojecten, Delft University of Technology, Delft.
Koutamanis, A.: 2007 (forthcoming), Use of a mobile architectural guide, CAAD Futures 2007, Springer, Sydney.
Marr, D.: 1982, Computer vision, W.H. Freeman, San Francisco.
Nelson, R. C.: 1997, From visual homing to object recognition, in Y. Aloimonos (ed), Visual navigation, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 218-250.
Nomura, Y. and Kishimoto, T.: 2005, Visualization of tourists' behavior and activity using GPS and GIS in Kamakura City, in A. Bhatt (ed), CAADRIA 2005. Proceedings of the 10th Conference on CAAD Research in Asia, CAADRIA, Seoul, pp. 320-327.
Persson, P., Espinoza, F., Fagerberg, P., Sandin, A. and Cöster, R.: 2002, GeoNotes: A location-based information system for public spaces, in K. Höök, D. Benyon and A. Munro (eds), Designing information spaces: The social navigation approach, Springer, London.
Riseman, E. M., Hanson, A. R., Beveridge, J. R., Kumar, R. and Sawhney, H.: 1997, Landmark-based navigation and the acquisition of environmental models, in Y. Aloimonos (ed), Visual navigation, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 317-374.
Robert, L., Zeller, C., Faugeras, O. and Hébert, M.: 1997, Applications of nonmetric vision to some visually guided robotic tasks, in Y. Aloimonos (ed), Visual navigation, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 89-134.
Shen, Y.-T. and Teng, T.-S.: 2005, Personal mobile device for situated interaction, in A. Bhatt (ed), CAADRIA 2005. Proceedings of the 10th Conference on CAAD Research in Asia, CAADRIA, Seoul, pp. 382-387.
Stellingwerff, M. C.: 2005, Virtual context, Delft University Press, Delft.
Waltz, D.: 1975, Understanding line drawings of scenes with shadows, in P. H. Winston (ed), The psychology of computer vision, McGraw-Hill, New York.