BRILL

Seeing and Perceiving 25 (2012) 303–338

brill.nl/sp

Shading, a View from the Inside

Jan J. Koenderink [1], **Andrea van Doorn** [2] **and Sylvia Pont** [2]

[1] EEMCS, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands; Experimental Psychology, University of Leuven, Leuven, Belgium, and The Flemish Academic Centre for Science and the Arts, Brussels, Belgium

[2] Industrial Design, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands

Received 24 June 2010; accepted 13 June 2011

Abstract

Shape from shading arose from artistic practice, and later experimental psychology, but its formal structure has only been established recently by computer vision. Some of its algorithms have led to useful applications. Psychology has reversely borrowed these formalisms in attempts to come to grips with shading as a depth cue. Results have been less than spectacular. The reason might well be that these formalisms are all based on Euclidean geometry and physics (radiometry), which, are the right tools in third person accounts, but have little relevance to first person accounts, and thus are biologically (and consequently psychologically) of minor interest. We propose a formal theory of the shading cue in the first person account, 'a view from the inside'. Such a perspective is also required for autonomous robots in AI. This formalism cannot be based on Euclidean geometry, nor on radiometry, but on the structure of pictorial space, and the structure of brightness space. The formalism, though different in kind, has a simple relation to the computer vision accounts. It has great robustness, is free from calibration issues, and allows purely local shape inferences. It is especially suited to biological (and thus AI) implementation. We consider a number of predictions and confront them with available empirical evidence.

Keywords

Shape from shading, shading cue, depth cues, pictorial vision, pictorial space

1. Introduction

We consider the 'Shape From Shading' (SFS) problem of computer vision (CV), and the shading shape cue of the experimental psychology of perception. The former belongs to the third person account of perceptual accomplishments, whereas the latter is interpreted in terms of a first person account of perceptual awareness. The formalism presented in this paper applies to the human condition, as well as to that of autonomous, intelligent, monocular robots.

DOI:10.1163/187847511X590923

1.1. Background

'Shape From Shading' is one of the formally best understood pictorial shape cues. It has especially been developed in the context of computer vision (Horn, 1970; Forsyth and Ponce, 2002). Virtually all formal development has been done in the context of a rather simplified model of radiometry, the major assumptions being:

- configurations of interest are smooth surfaces of opaque objects. The surface properties are uniform;
- the radiometry is fully described in the single bounce approximation, that is to say, multiple scattering is ignored;
- the bidirectional reflectance distribution function is constant. These are the so called Lambertian surfaces (Lambert, 1760). Without loss of generality one assumes unit albedo;
- each surface element can 'see' all of the source, thus vignetting is ignored;
- the primary radiators are located at a large distance as compared to the extent of the scene.

This has quite a number of important consequences. An observer looking at the scene receives a certain radiance of beams that have been scattered to the eye from surface elements in the scene (Adelson and Bergen, 1991). Given these assumptions this radiance is proportional to the irradiance of the surface. The viewing geometry plays no role, only the illumination geometry does. The light field can be summarized through the light vector (Gershun, 1936), a uniform field over the volume of the scene, thus, it can be further summarized through a single direction. Perhaps somewhat perversely, but conveniently, one uses the direction toward the source. Then the 'shading' is given through 'Lambert's cosine law' (Lambert, 1760), it is proportional to the cosine of the angle subtended by the outward normal of the local surface element and the light direction. Usually no radiometric calibration is available, thus one uses 'image intensities' (pixel values) multiplied by a factor that renders the maximum value less or equal than one in order to obtain the values of the cosine. As a consequence one is left with a purely geometrical problem. The observation yields a scalar field of the cosine of the outward surface normal with a fixed direction. The task is to find the surface and the direction. A commonly encountered variation assumes the direction 'given'.

The problem is an interesting one from a formal perspective, because it has no unique solution. (As typical for the sciences, and in contradistinction to pure mathematics, the problem of whether there exists a solution at all does not even arise in the case of an actual observation.) One finds an infinite family of solutions (the 'bas-relief ambiguity', Belhumeur, Kriegman and Yuille, 1999), much to the chagrin of the CV society. This again gives rise to a large number of methods (Zhang *et al.*, 1999) that we will not consider here, but that are of obvious interest to CV applications.

The methods used in the solution invariably involve Euclidean differential geometry (Coxeter, 1989). This is an immediate consequence of Lambert's cosine law, which involves the surface normal, an intrinsically Euclidean object.

1.2. The View from the Inside

The CV formalism for SFS is generally considered to be the natural platform from which to start psychological investigations of the 'shading cue' (Palmer, 1999; Poggio, 1984). This is indeed the motivation to have been a participant ourselves in the development of the current framework. Perhaps unfortunately, the advances of psychology in this area have been less than spectacular, most of the current understanding was around in the first half of the twentieth century (Metzger, 1953), quite remarkable in view of the obvious more recent progress in CV.

The reason, we believe, is obvious enough, though perhaps only in retrospect. The human observer is quite unlike any of the CV algorithms in that the observer cannot draw on established physics (Berkeley, 1709). The observer (perhaps over evolutionary time spans) has to develop an interface that is efficacious (von Helmholtz, 1860). Efficacy is what counts evolutionary, 'veridicality' as such being irrelevant (Riedl, 1975; Lorenz, 1977; Tinbergen, 1951, 1975; von Uexküll, 1921) (survival or fitness, and the pursuit of truth are quite different goals). The observer can hardly be supposed to use Euclidean geometry, given the fact that visual space is necessarily non-Euclidean because 'depth' and 'visual field' dimensions have very different ontologies. Neither can the observer be supposed to use regular radiometry, indeed every indication we know would suggest that observers do not.

From a phenomenological perspective, we are interested in the microgenesis (Brown, 2000; Rosenthal, 2004) of immediate visual awareness. Immediate visual awareness is a sequence of 'presentations' that simply happen to an observer. They are pre-cognitive, and the process of microgenesis is subconscious. Microgenesis generates presentations at a rate of about a dozen a second. Perceptions are cognitions based on maybe a dozen of presentations in a 'specious moment'. We are not involved with reflective thought in this paper, our aim is the nature of the microgenetic process. Since presentations happen to you, you have no voluntary control over them.

For the purposes of psychology the CV algorithms are to be considered sophisticated stimulus descriptions. They are not candidates for descriptions of the microgenetic processes, because they refer to entities and structures that are not available to the observer. Consequently, the use of CV algorithms in a psychological context implies that the microgenetic process is treated as a black box. This equally applies to artificial intelligence (AI). An automaton programmed from the outside, and sent out into the world is not an 'intelligence', but a machine. For the creature to be truly intelligent (as distinct from efficacious by design) it needs to develop a 'view from the inside' (see below).

Thus, arises the need to develop a formal theory of SFS 'as viewed from the inside' as it were, what is to say, from the observer's perspective. The result of

such 'computations' would be pictorial reliefs instead of surfaces in Euclidean three-space, and the computations will be done on the basis of brightness contrast gradients, rather than radiance. In such a setting it is inconsistent to use Lambert's cosine law directly, since it involves entities (the Euclidean outward normal) that play no role in the observer's mental framework.

If one takes these notions seriously, and it is our conviction that one has no alternative, it implies that — for purposes of psychology and AI — one needs to construct a completely novel formalism from the ground up. This is the aim of our present exercise. In order to develop such a formal theory we need to draw on existing formal accounts of pictorial space and subjective radiometry. We will introduce these subjects — in summary fashion — first, then proceed with the development of the theory of the shading cue.

The issue of the ambiguity of possible inferences is an important one in framing methods, because what is to be considered a useful result depends upon the potential use. We consider two common cases. Either the observer needs to come up with a definite result, or the result is made available as a constraint on the process as a whole. In the former case the observer needs to stick its neck out and has to rely on guesses, for better or worse. These can range all the way from a random choice from the set of inferences not ruled out by the observation, to a full fledged Bayesian optimal choice. In the latter case the result of the shading cue will be a constraint on inferences enabled by other cues. This is probably the generic case (Erens, Kappers and Koenderink, 1993). The constraint should preferably be of universal applicability in any setting. We consider both cases.

2. Preliminary Considerations

We consider three concepts that will be crucial in the development of the alternative SFS-formalism.

The first is the concept of perception as a user interface. This is predominantly a conceptual matter, although its acceptance has important consequences. The idea is common enough in biology (the field of ethology), but, with notable exceptions (Hoffman, 2008), it is not considered seriously in psychology. Although indeed largely conceptual, we believe that forthcoming empirical data is going to force this notion on the field. Collecting empirical evidence is only starting, due to the different kind of research questions that have to be considered (Koenderink, van Doorn and Todd, 2009). Thus, the evidence is certainly not to be found in the mainstream. In the case of AI the necessity of the interface view of perception is obvious enough, the alternative would be a preprogrammed machine.

The concept is important in the present context because it implies that the entities playing a role in microgenesis need not be representations of physical entities. As the icon on your laptop screen does not represent any electronics or systems programs, but stands for an entity of your reflective thought (a text maybe), the ele-

ments of immediate visual awareness stand not for physical properties of the scene in front of you. They are 'mental paint' (*qualia*).

The second is the formal theory of subjective radiometry. Here is much empirical material to be mined (perhaps starting with Fechner's psychophysical law, Fechner, 1860). A formal description that covers the area is lacking though. We discuss an attempt that can at least be drawn upon for the present purpose. It can be shown to be the optimal choice in the absence of prior knowledge, thus it is also the appropriate choice for the AI case.

The third is a formal theory of pictorial space. We use 'pictorial' in order to stress the case of the monocular, static observer. Otherwise 'pictorial space' is not different from 'visual space'. Here exists a formal theory with excellent, quantitative predicting power and a large corpus of empirical data (Koenderink and van Doorn, 2008). At least for the present purposes, the formal geometry of pictorial space may be considered to have been established. In retrospect the structure can be deduced from general first principles that equally apply to autonomous, monocular robots, thus, the choice is also an apt one for AI.

We will not discuss the empirical foundations here, as they can be found in the above reference. But the reader is advised to take the geometrical structure that is the backbone of the formalism developed in this paper as firmly established rather than merely hypothetical.

2.1. Perception as a User Interface

We will use 'presentation' to denote the visual awareness that you experience when you open your eyes in front of a scene. It differs from 'perception' in that it happens to you, like a sneeze, whereas perceptions are reflective thoughts, something you do. Since presentations are not thoughts they are not part of cognition, but perhaps can be considered to be at the fringe of it, supplying thoughts with their substance so to speak (Albertazzi, 2008; Poli, 2001; Searle, 1983). To some extent one may study the microgenesis of presentations, but the microgenetic process proper is not part of awareness (Wohlfahrt, 1932). The Gestalts from early twentieth century psychology (Metzger, 1953) are part of presentations, they simply occur, there is nothing you can do about them. The 'releasers' and 'imprintings' described by the ethologists (Lorenz, 1977; Tinbergen, 1951, 1975; von Uexküll, 1921) are likewise part of the presentations.

The conclusions of the ethologists are of much interest, because they provide a biological, evolutionary basis for the structure of presentations. Presentations are in no way 'representations of the scene in front of you', and the issue of their 'veridicality' is void. Organisms or autonomous robots do not attempt to invent physics (the matter of veridicality), nor does the notion of 'representation' make sense for them. Representation in the first-person account implies extra sensory perception, thus is not a scientifically acceptable notion. Organisms simply develop efficacious user interfaces, this is what drives evolution (Hoffman, 2008; Riedl, 1975; von Uexküll, 1921). An interface need neither be veridical, nor need it be a

representation. Indeed, the most useful interfaces are neither of these. On the contrary, they shield the user from the real world (which only exists in the third-person account), just like the icons on your computer desktop. Most of us do not know what happens in the innards of our laptop when we drag a square icon (representing a file) on another icon (representing the 'trash'), nor do we have any desire or necessity to know. That would stand in the way of efficacious interaction. It is only the interface that is relevant, to most of us our computer is the interface, rather than the electronics or systems programs. For the autonomous robots of AI the case is not different.

Interfaces are constructions of the observer and they are by their very nature idiosyncratic (Brown, 2000). Since we interact with each other *via* the physical world, there are no problems with communication. It is only that our 'mental paints' may be quite distinct, the 'problem of *qualia*' (Albertazzi, 2008; Block, 2003; Brentano, 1995). Generic methods in experimental psychology stress objectivity and downplay first person accounts, which is why such (as we believe very common) idiosyncrasies are not that often being reported. It requires a small paradigm shift to allow them as scientifically valid facts, but once one accepts this one encounters them abundantly (Koenderink *et al.*, 2001; Koenderink, van Doorn and Todd, 2009). Such reports are not received with enthusiasm by the mainstream literature of experimental psychology though, since they go against the grain of what is considered scientifically acceptable. This is one major reason why there exists scarce documentation.

The upshot is that the shading cue is processed in the microgenesis of presentations, but that it by no means need to reflect the standard formalism of CV, nor need it result in reconstructions of the scene in front of you. Shading gives rise to pictorial relief, which needs not be identical to the Euclidean differential geometry of a surface in the scene in front of you. This is the human condition, but it applies equally to the AI case.

2.2. Subjective Radiometry

The subjective entity that roughly correlates with radiance is brightness. These are categorically distinct. Radiance (Born and Wolf, 1999) is non-negative, and is measured in some conventional unit, say photons per surface area, per solid angle, per second, and per photon energy interval. In contradistinction, brightness has no natural origin, nor does it have a natural unit. Thus, whereas radiance may be parameterized by the Euclidean half-line, brightness is naturally parameterized through the affine line (Koenderink and van Doorn, 2002). Fechner's psychophysical law expresses this: $\Phi = \log(N/N_0)$, where Φ denotes the brightness, N the radiance, and N_0 is some arbitrary unit. If $N(x, y)$ denotes a 'picture' (the retinal illuminance will be proportional to it), then $\Phi(x, y)$ denotes the corresponding image. We will not use brightness as a point property, but consider only brightness contrast gradients as observables. This reflects the properties of the front-end re-

ceptive fields and basic psychophysics. However, its use as a formal parameter (as in the expression below) leads to greater formal transparency.

It is an empirical observation (Koenderink *et al.*, 2001) that the images $\Psi(u, v)$, related to the image $\Phi(x, y)$ by the transformation

$$\begin{pmatrix} u \\ v \\ \Psi \\ 1 \end{pmatrix} = \begin{pmatrix} h\cos\mu & -h\sin\mu & 0 & t_x \\ h\sin\mu & h\cos\mu & 0 & t_y \\ s_x & s_y & g & t_\Phi \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ \Phi \\ 1 \end{pmatrix} \tag{1}$$

with $h, g > 0$, are in most respects to be considered equivalent ('same picture') to the human observer. This is why people believe to all watch the same TV show, even when their TV pictures differ a lot, largely described through these equations. Most people only notice this when they turn on a cheap TV set in a motel room that has unfamiliar (perhaps due to a previous guest or simply aging of the electronics) settings. Virtually all of these parameters have control knobs, although typically hidden away in the TV box somewhere.

The first two rows in equation (1) represent simply Euclidean similarities in the picture (and image) plane, but the third row defines transformations that involve brightness:

$$\Psi = s_x x + s_y y + g\Phi + t_\Phi, \tag{2}$$

it is the really interesting part. These are the transformations that photographers used to do in the darkroom (Adams, 1950), and are nowadays standard tools in Photoshop, all considered part of 'straight photography'. A huge transformation will be accepted, whereas a minor one will not even be noticed by most observers.

Equation (1) describes the group of similarities of the three-dimensional Cayley–Klein space (Cayley, 1859; Klein, 1871, 1893, 1928) (like the familiar 'space you move in' is) with two Euclidean and one isotropic dimension. Thus, it is only slightly different from Euclidean three-space that might be more familar to you. It is a well-known space (Sachs, 1987, 1990; Strubecker, 1941, 1942, 1943, 1945; Yaglom, 1979). In our application we will only use the subgroup defined by $h = t_x = t_y = 0$, because we have no reason to transform the image plane. In the resulting subgroup the parameters $\{s_x, s_y\}$ describe an isotropic rotation, t_Φ an isotropic translation (both congruencies) and g a scaling of isotropic angles (similarity of the second kind). In most cases we will also set $s_x, s_y = 0$, as these parameters are rarely used in photographic image processing, mainly for things as 'edge darkening' (Adams, 1950), which have no generic interest.

We will consider some image $\Phi(x, y)$ as the 'observable', the input to the shape from shading process. This is quite a deviation from the CV setting (Forsyth and Ponce, 2002), because it allows for arbitrary 'gamma corrections' (parameter g), and 'contrast changes' (parameter t_Φ). The input data thus being more ambiguous, one expects the problem to be more difficult, and the group of resulting ambiguities to be larger.

Although this structure appears to account for many of the facts of human perception, it can actually be derived from first principles (Koenderink and van Doorn, 2002), thus, it is the obvious choice for the autonomous robot in AI.

2.3. The Structure of Pictorial Space

In this exercise we will be mainly occupied with pictorial reliefs, that are surfaces in pictorial space. We will parameterize points in pictorial space by $\{x, y, w\}$, where $\{x, y\}$ parameterize the location in the picture plane (or in the visual field, we will not distinguish these here). The coordinate w is 'pictorial depth', which is not specified optically, but is a mental entity that results from the microgenesis of presentations. A pictorial relief is given as a surface $w(x, y)$ in pictorial space (Koenderink and van Doorn, 2003a). A physical correlate would be the egocentric distance (we will simply refer to it as 'distance') z (say). We do not a priori expect there to exist a fixed monotonic (or even any functional) relation between depth and distance (so-called 'distance function') though.

Distance is a non-negative entity, with some conventional unit. Depth has neither an origin (the eye is not in pictorial space), nor a natural unit.

Absolute depth is meaningless in pictorial perception and cannot be measured experimentally. It is possible to measure the depth gradient $\nabla w = \{w_x, w_y\}$ (where the subscripts denote differentiation to that coordinate), that is the surface attitude. A major finding is that transformations of the type

$$\nabla w'(x, y) = \{s_x, s_y\} + g\nabla w(x, y), \tag{3}$$

with $g > 0$, otherwise arbitrary and s_x, s_y arbitrary account in detail for the differences between observers in the same task, and for differences over time for a single observer for the same task (Koenderink *et al.*, 2001). Such idiosyncratic differences can be surprisingly large. These ambiguities have already been described in the visual arts at the end of the nineteenth century (Hildebrand, 1893), though psychology has taken little notice of this.

This suggests that pictorial space — like the space of brightness images — is a Cayley–Klein space (Cayley, 1859; Klein, 1871, 1893, 1928) with two Euclidean (the picture plane) and one isotropic (depth) dimension. (Taking the gradient of equation (2) leads to the form of equation (3).) This geometry is very similar to Euclidean geometry (which is also one of the twenty-seven three-dimensional Cayley–Klein spaces). It can be (but of course need not be) understood as an approximation to Euclidean space when one Cartesian dimension is considered infinitesimal. This geometry can be derived from first principles (Koenderink and van Doorn, 2008) on the assumption that vision is invariant with respect to arbitrary rotation–dilations about the vantage point. This is intuitive, since Lilliput looks no different from Brobdignac to the proper inhabitants (Swift, 1906). These are the transformations that leave the optical structure available at the eye invariant, thus it is a very general condition. It equally applies to the human condition as to the autonomous robot from AI.

The major group of ambiguities in the formal theories of SFS in CV (Belhumeur, Kriegman and Yuille, 1999; Forsyth and Ponce, 2002) is exactly described by equation (3), no doubt for the aforementioned reason. The term $\{s_x, s_y\}$ is conventionally denoted 'added plane' and the factor g 'depth scaling'. Of course CV applies this not to depth, but to recovered, or estimated, distance.

3. The SFS Problem as Viewed from the Inside

The transformations of singly isotropic space never mix the depths along distinct visual rays (equation (3)). Thus, pictorial space may be conceived of as a fiber bundle (Steenrod, 1951) with base space the Euclidean plane (the picture plane), and fibers the visual rays (fixed $\{x, y\}$, free w). The canonical projection simply forgets the depth. A pictorial relief is a cross section of this fiber bundle, a smooth assignment of depth values to the fibers (Fig. 1). Thus, the task of the microgenetic process in shape from shading is to assign these values, moving depth estimates along the visual rays as beads on strings (Fig. 2). It does so on the basis of cues, in this case the shading cue.

Notice that this concept of pictorial space as a fiber space is actually more apt, but also formally simpler, than the notion of Euclidean three-space. It is in many respect (formally) similar to David Marr's notion of a 'two-and-a-half dimensional sketch' (Marr, 1982). The 'half' dimension is the isotropic dimension, that is the depth dimension. Of course Marr never was explicit about the meaning of a 'half' dimension. Here we have a formal description of it.

In the user interface model it is evident that the microgenetic process also has the task to *define* the cues. Cues are not somehow 'given', all that may be considered 'given' is (necessarily meaningless) recorded optical structure.

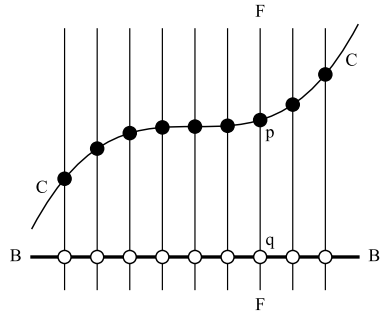

Figure 1. The notion of a fiber bundle. The base space BB represents the visual field, whereas the fibers (such as FF) are 'visual rays', that is the depth dimension. A cross section like CC is a 'pictorial relief', each fiber has been assigned a depth. The canonical projection (or bundle projection) maps points of the fiber (like p) on the base space (in this case q). In microgenesis the mind shifts depth values (like point p) along the visual rays like beads on a string. Points never leave the fiber they belong to. (Of course, the actual base space — the visual field — is two dimensional, the figure merely suggests the essential formal relations.)

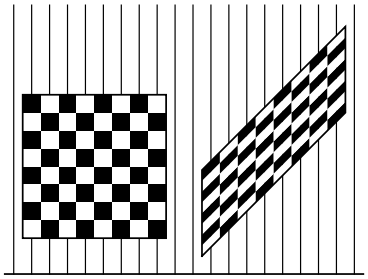

Figure 2. A transformation in the fiber bundle, fibers the verticals, base space horizontal. We have applied a shift along the base space in order to show the original (left) and transformed configuration (right) separately. Notice that points shift only along the fibers, thus, this is not a Euclidean rotation, although it does manage to tilt the apparent frontoparallel. The transformation is exactly that described by equation (3). It is typical for what is found in the psychophysics of pictorial space. (The checkerboard is only a convenient visualization here: this figure shows only a planar cut through visual space, thus the 'visual image' is linear, no room for a checkerboard!)

The optical structures that might be of interest in shape from shading are brightness gradients. Zero gradient implies a uniform region, which could be interpreted as the result of an illuminated plane of arbitrary spatial attitude. Such a plane can be removed with a transformation of the type (3), it is of no interest to the microgenetic process. Any gradient is at least potentially a shading cue. Of course it might be due to other causes too. If the process takes it for shading it sticks out its neck so to speak, it is a leap in the dark. Only the result will tell whether the guess was a fortunate one or not. If not, a fresh assignment has to be made (e.g., not shading, but a reflectance variation). The microgenetic process is naturally iterative and recursive. Here we only consider the case that the shading cue assignment is a useful one.

The assignment need not to be correct for it to be useful. For instance, it is often useful to 'see' three-dimensional scenes in photographs whereas these are actually flat objects covered with pigments. This latter, technically to be considered 'veridical', meaning is more useful if you want to hang the picture on the wall, or use it as a mouse mat. On the other hand, the assignment is useful if you want to use the contents of the picture as forensic evidence. The choice is entirely up to you.

A brightness gradient may be taken as the cue for a change of tangent plane of the pictorial relief with respect to some 'light direction'. Of course, light direction can not be taken in the radiometric sense here. The light direction matters. For instance, if a cylindrical surface is illuminated from a direction that is coplanar with the cylinder axis no shading will result, whereas if it is illuminated from a direction that is skew to the axis it will. Thus, if the surface is known to be a cylinder, the observation of a gradient rules out certain light directions.

In the pictorial space descriptions surface normals are useless (any visual ray being perpendicular to any surface element, thus all 'normals' being mutually parallel), but a good alternative is to use depth gradients. Depth gradients define tangent planes, which is what really counts. (Even in the CV setting the normal serves pri-

marily to indicate the tangent plane as the causally effective factor, in Euclidean space the relation is immediate.) Since depth gradients live in the two-dimensional tangent space of the picture plane, the alternative for a light direction has to be a two-dimensional vector in that plane too.

In setting up a formalism for the shading cue, one may take some inspiration from the formalism of the external view. Consider a surface described by the distance function $z(x, y) = \frac{1}{2}(z_{xx}x^2 + 2z_{xy}xy + z_{yy}y^2) + \mathrm{O}[x, y]^3$, thus, the tangent plane at $x = y = 0$ is frontoparallel, and the relief considered shallow (thus, $\sqrt{z_x^2 + z_y^2} \ll 1$). Let the light vector be $\{\sin\vartheta, 0, \cos\vartheta\}$, in the xz-plane. Then Lambert's cosine law yields the illuminance $I(x, y) = \cos\vartheta - (xz_{xx} + yz_{xy})\sin\vartheta$, thus the illuminance gradient $\nabla I|_{\{0,0\}} = -\sin\vartheta\{z_{xx}, z_{xy}\}$. The expression that finally appears, and that may serve as a heuristic for the formalism of the view from the inside is

$$\frac{\nabla I|_{\{0,0\}}}{I_0} \approx -\tan\vartheta\{z_{xx}, z_{xy}\}, \tag{4}$$

where $I_0 = I(0, 0) = \cos\vartheta$ and $(\nabla I|_{\{0,0\}})/I_0$ is the 'contrast gradient'. This is a pretty relation in its own right, but it is only of use to an external observer.

Now for the view from the inside. Suppose the light direction is the vector $\mathbf{a}$, say, and the local depth gradient ∇w. Both are vector fields in the visual field, the depth dimension is not involved. Then the simplest expression with the qualitatively correct properties to represent a 'shading' is $\mathbf{a} \cdot \nabla w$. If we agree on $\|\mathbf{a}\| = 1$ (Euclidean length is well defined in the picture plane!) we obtain a number on the real line $\{-\infty, +\infty\}$. Thus, we set

$$\Phi(x, y) = \mathbf{a} \cdot \nabla w(x, y), \tag{5}$$

as the alternative to Lambert's cosine law.

Further on, we will show that this can be interpreted as a low order ($\vartheta \ll 0$) approximation (Pentland, 1988) to the Euclidean–radiometric case (equation (4)). For some people, this may put their minds at rest. However, this is not how we will view it at all! Here we are in an ontologically completely different universe. Equation (5) is essentially an independent construction, fully in the spirit of the 'perception as user interface' concept. It is a bold hypothesis, rather than some approximation. It makes good sense from a general, formal point of view (being much more general than the Euclidean–radiometrical setting), but it could be completely wrong as a biological or psychological theory. Thus, equation (5) is to be considered to be a speculative hypothesis (Fig. 3).

From equation (5) we find the brightness gradient:

$$\nabla\Phi(x, y) = \mathrm{H}_{w(x,y)}\mathbf{a}, \tag{6}$$

where H denotes the Hessian of the pictorial relief. If the parameters $\{s_x, s_y\}$ in equation (2) are zero (they are typically very small) it turns out that (apart from the light direction) the crucial entity is the Hessian of the pictorial relief up to a factor. An arbitrary factor is no problem anyway, since the relief itself (equation (3)) is

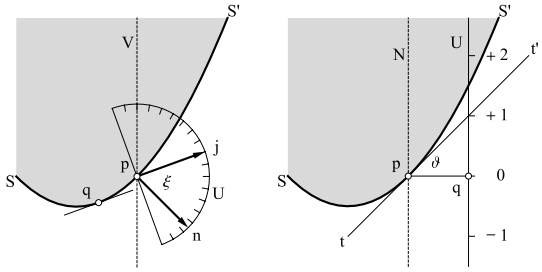

Figure 3. Comparison between the views from the outside (left) and from the inside (right). In the left-hand figure the cosine of the angle ξ subtended by the surface normal n and the light vector j determines the illuminance at p on the visual ray V. The angle ξ is read on the protractor U, connected at p, oriented towards j. At the left of point q the surface SS' is not illuminated due to the attached shadow. In the right-hand figure the illumination is towards the left, shadows are meaningless. The isotropic angle ϑ determines the illumination, it can be read of on the scale U, which serves as a protractor in this geometry (U is connected to p and at unit distance from it). This is formally just Lambert's cosine law (for ϑ mimics $\sin(\frac{\pi}{2} - \xi) = \cos\xi$ in the Euclidean case), the difference is in the angle measure, which is hyperbolic (not periodic), and the non-Euclidean trigonometry ($\cos\vartheta = 1$, $\sin\vartheta = \vartheta$). Notice that both N and U are normal to the surface SS' in this geometry, isotropic normals are useless for most purposes.

only of interest up to a factor. The Hessian concerns only the curvature, the spatial attitude of the relief again being undefined (the result of equation (3)).

Thus, one ends up with an alternative to Lambert's cosine law that uses only parameters of interest to the microgenetic process. We consider the relation to the generic CV formalism later.

Although this might appear like a mere crummy version of the real thing, it has the advantage of being very robust. Both the observables and the estimates are only defined up to a large group of ambiguities, thus calibration issues hardly arise.

Written out explicitly equation (6) is

$$\Phi_x = A(a_x w_{xx} + a_y w_{xy}), \tag{7}$$

$$\Phi_y = A(a_x w_{xy} + a_y w_{yy}), \tag{8}$$

for some arbitrary $A > 0$, or, equivalently,

$$\Phi_y a_x w_{xx} + (\Phi_y a_y - \Phi_x a_x) w_{xy} - \Phi_x a_y w_{yy} = 0, \tag{9}$$

a bilinear, homogeneous equation in $\{w_{xx}, w_{xy}, w_{yy}\}$ and $\{a_x, a_y\}$.

It is already possible to draw some initial conclusions. Suppose the light direction is (somehow) known. Then we may take $a_x = 1$, $a_y = 0$, without loss of generality. From equations (7) and (8) we find that we obtain w_{xx} and w_{xy} up to a common scaling, whereas w_{yy} remains indeterminate. Thus, along a field line of the flow, we obtain a strip of pictorial surface. Due to the uniformity of the light field the strip has zero geodesic curvature, and the solution yields its normal curvature (w_{xx}) and twist (w_{xy}) up to a common constant.

Suppose we know the surface to be umbilical. Without loss of generality, we set $w_{xx} = w_{yy} = 1$, $w_{xy} = 0$, and find that the light direction equals the brightness gradient direction.

The light direction evidently constrains possible shapes and *vice versa*. Thus, the shading cue leads to an infinite family of solutions. The microgenetic process has two options (apart from giving up, which is no option) open, either to just guess at the lacking information, or to use additional cue information to narrow down the possibilities. Judging from the literature (Palmer, 1999; Ramachandran, 1988), the former option is often adopted. Observers report to see either a concave or a convex umbilical shape and are consequently able to judge the light flow direction. This is quite an extreme move, because in reality *any quadric shape qualifies as a possible solution*.

Why are these perfectly good solutions never reported in practice? We believe that the reason is that no experiment addresses the shading cue proper, in all cases there are additional cues that appear to force the issue (see below).

4. Light Flow as Viewed from the Inside

Although we assume a uniform light direction throughout space, this is not immediately evident from the optical structure. Assuming opaque, Lambertian objects, with — for the moment — smooth surfaces, the only causally effective component of the light vector is its normal component at the location of material surfaces. The result is the shading that is observed. In the conventional SFS setting the tangential component is causally ineffective.

This changes if one assumes opaque surfaces that have relief modulations on the mesoscale. Here 'mesoscale' is the region between the megascale at which objects have 'shapes', and the microscale which influences the optical properties of the surface, but remains unresolved in the image. Such relief modulations give rise to 'texture', a modulation pattern in the image intensity at a scale below that of relevant shape variations. The nature of this texture depends upon the illumination geometry (Chantler, Russell and Linnett, 1994; Koenderink and Pont, 2003), a fact that distinguishes it from 'wallpaper texture' due to albedo variations. In this paper we assume constant albedo throughout. The texture is easily shown to depend upon the tangential component of the light vector. This component, on the illuminated surface of objects, we call '(surface) illumination flow', projected in the image we speak of '(image) light flow'. (In the interest of succinctness we will often omit 'surface' or 'image'.) The orientation of the flow (direction up to 180° ambiguity) is imprinted on the statistical structure of the texture. The gradient structure tensor allows the estimation of image light flow, human observers are sensitive to this and estimate orientation with an accuracy of better than 10° (Koenderink, van Doorn and Pont, 2004). This aspect of the illumination induced optical structure is ignored in the classical SFS setting. It is often of relevance in actual situations though.

Let $\mathbf{i}$ be the light direction, and $\mathbf{i}_\perp$ its projection in the visual field, renormalized to unit magnitude. The lost inclination (captured in $\mathbf{i}_\parallel$, the component of $\mathbf{i}$ along the visual direction) is ε, say. Let the distance gradient of the surface be $\mathbf{g}(\mathbf{r})$, where $\mathbf{r}$ denotes the location in the visual field. Then, to first-order in the slope and the inclination, the light flow is

$$\mathbf{f}(\mathbf{r}) = \mathbf{i}_\perp + \varepsilon \mathbf{g}(\mathbf{r}), \tag{10}$$

where $\mathbf{f}$ is, to first-order, of unit magnitude. (That is to say, we assume $\varepsilon^2 = 0$, $\varepsilon \neq 0$, thus, ε is a 'nil-square infinitesimal'.) Thus, the flow direction depends upon the local spatial attitude of the surface. The exact expressions are rather more complicated than this first-order approximation of course. This is the Euclidean analysis, that is to say, the 'outside view' of the matter. We use it as a heuristic in the construction of a 'view from the inside'.

The Euclidean analysis is purely objective, and essentially trivial. It is quite another matter to frame a speculative theory that purports to capture the view from the inside. The fact that we framed the Euclidean case in terms of a linear approximation is not arbitrary, for we believe it *a priori* likely that the view from the inside will be most aptly described by way of a linear theory.

As viewed from the inside, the field of flow directions $\mathbf{f}(\mathbf{r})$ is an observable, whereas both the component of the light vector in the viewing direction and the gradient $\mathbf{g}(\mathbf{r})$ are not observable, but, when necessary, need to be constructed. In order to be able to do so we use equation (10), interpreted from the inside perspective. As is well known, the flow direction is crucial in SFS. Perhaps less appreciated is that the spatial variation of the flow direction is a shape cue in its own right. From the Euclidean analysis you have

$$\mathbf{f}(\mathbf{r} + \delta\mathbf{r}) = \mathbf{f}(\mathbf{r}) + \varepsilon H \delta\mathbf{r}, \tag{11}$$

where H denotes the Hessian of the distance. Locally, this is best expressed in terms of two differential invariants, the rate of turn of the flow direction for a unit progression in the flow direction, and the rate of turn of the flow direction for a unit progression orthogonal to the flow direction. The former leads to a curvature of the flow lines and will be denoted 'swerve' σ, the latter leads to a progressive divergence of the flow lines and will be called the 'splay' τ. One has

$$\sigma = \varepsilon\, z_{xy}, \tag{12}$$

$$\tau = \varepsilon\, z_{yy}, \tag{13}$$

where x denotes a coordinate in the flow direction, and y a coordinate orthogonal to it (Fig. 4). By construction, splay and swerve are differential invariants.

This then is the flow of light account for the view from the inside: The flow is observed, although usually with a 180° ambiguity. Extraneous cues may in some cases lift the ambiguity. The mechanism can be a non-linear receptive field, implementing the gradient structure tensor

$$S = \begin{pmatrix} \langle I_x I_x \rangle_{\mathcal{R}} & \langle I_x I_y \rangle_{\mathcal{R}} \\ \langle I_y I_x \rangle_{\mathcal{R}} & \langle I_y I_y \rangle_{\mathcal{R}} \end{pmatrix}, \tag{14}$$

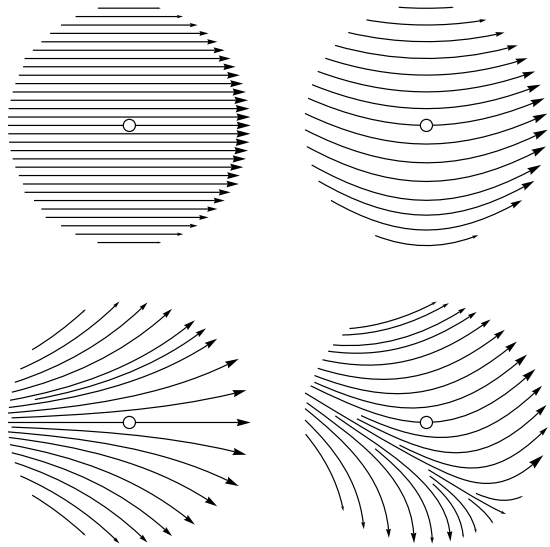

Figure 4. Some flow fields for a surface $\frac{1}{2}(z_{xx}x^2 + 2z_{xy}xy + z_{yy}y^2)$, where x denotes the light direction (horizontal in the figure). At top left one has $z_{xx} \neq 0$, $z_{xy}, z_{yy} = 0$ (neither swerve, nor splay), at top right one has $z_{xy} \neq 0$, $z_{xx}, z_{yy} = 0$ (pure swerve), at bottom left one has $z_{yy} \neq 0$, z_{xx}, $z_{xy} = 0$ (pure splay). At bottom right a general case with both swerve and splay.

where the averaging is over a region $\mathcal{R}$ that contains a sufficient number of texture elements. The flow is understood as a flow field over the pictorial relief. The other observables are the local variation of the flow direction, the swerve and splay. They are understood as being proportional to the curvature of the relief at right angles to the flow, and the mixed derivative. Thus, the view from the inside is based on the (speculative) relations

$$w_{xy} = \frac{\sigma}{\mu}, \tag{15}$$

$$w_{yy} = \frac{\tau}{\mu}, \tag{16}$$

where μ is a parameter of the inner view that captures the influence of ε. The mechanism would be receptive fields that implement spatial derivatives of the flow direction, thus fields that are larger than those that serve to monitor the flow itself. Swerve and splay are of considerable interest, since the curvature of the surface at right angles to the flow does not affect the shading, and thus leads to problems for pure SFS inferences. Of course, it may not always be possible to observe either the flow direction, or the swerve and splay, for instance if the surface happens to be very smooth.

This ‘inside view’ is a simple one, that will make sense as seen from the outside when the observed surfaces are not slanted too much. In Fig. 5 we show the example of a sphere with arbitrarily high slants as one approaches the contour. It yields an idea of the utility of the simple inside view.

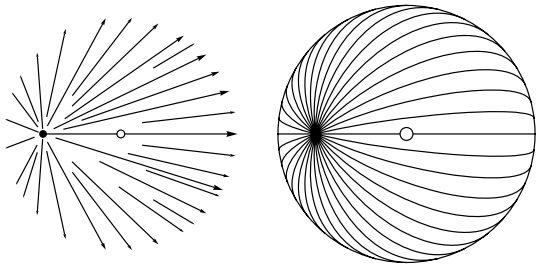

Figure 5. The example of a sphere. At right is the exact flow field. At left the field for the analogous case in the formalism captured by equation (11).

5. Inferences

Shape from shading, in the first person account, is often considered an inference on the basis of currently available optical structure and situational awareness (Minsky, 1974; Searle, 1983). An extreme view considers it even as 'inverse optics' (Poggio, 1984).

From a biological perspective 'inference' has perhaps too much of a cognitive connotation. Here we consider the microgenesis of perception to be pre-cognitive, somewhat close to the notions of the early Gestalt school. This is how 'inference' should be read.

In an AI setting the 'inferences' would be results of a first stage of default processing, prior to any logical or probabilistic reasoning.

5.1. Purely Local Inferences

So far, no one has proposed a purely local receptive field for pictorial shape on the basis of shading. The reason is evident from equations (7) and (8). Apparently the shape remains ambiguous even if the light flow direction is known. The same problem troubles the CV algorithms. One uses some global constraint in order to force a solution (Zhang *et al.*, 1999). For biological systems this appears to be little of an option.

One would prefer purely local inferences if at all possible. This would also be of interest for AI robotics designs. In order to make some progress we need a little digression on the topic of local shapes of quadric surfaces.

5.1.1. Quadrics

One advantage of the structure of pictorial space is that the differential geometry of surfaces is much simpler than it is in Euclidean space (Sachs, 1987, 1990; Strubecker, 1941, 1942, 1943, 1945; Yaglom, 1979). All one needs is the Hessian, whereas in Euclidean geometry one deals with nonlinear combinations of the Hessian and the gradient (Coxeter, 1989).

Consider the quadric

$$w(x, y) = \frac{1}{2}\left(a_{20}x^2 + 2a_{11}xy + a_{02}y^2\right). \tag{17}$$

We rewrite it in terms of the alternative coefficients $r = (a_{20} - a_{02})/2$, $s = a_{11}$ and $t = (a_{20} + a_{02})/2$:

$$w(x, y) = t\frac{x^2 + y^2}{2} + \left(r\frac{x^2 - y^2}{2} + sxy\right) \tag{18}$$

for the reason that $x^2 - y^2$ and $2xy$ transform as a pair under rotations (hence the parentheses), after all $x^2 - y^2$ equals $(x - y)(x + y)$, just another cross product, but for a 45° rotated coordinate system. Indeed, in rotated coordinates $\{u, v\}$ over an angle a, defined as $x = u\cos a - v\sin a$, $y = u\sin a + v\cos a$ you have

$$\begin{aligned} w(u, v) = t\frac{u^2 + v^2}{2} &+ (s\cos 2a - r\sin 2a)uv \\ &+ (r\cos 2a + s\sin 2a)\frac{u^2 - v^2}{2}. \end{aligned} \tag{19}$$

The term in uv will vanish if you set

$$a = \frac{1}{2}\arctan\frac{s}{r} \quad \text{or} \quad \cos 2a = \frac{r}{\sqrt{r^2 + s^2}}, \qquad \sin 2a = \frac{s}{\sqrt{r^2 + s^2}}, \tag{20}$$

then the quadric becomes

$$w(u, v) = t\frac{u^2 + v^2}{2} + \sqrt{r^2 + s^2}\frac{u^2 - v^2}{2}. \tag{21}$$

Thus, the principal curvatures are $\kappa_{1,2} = t \pm \sqrt{r^2 + s^2}$, whereas a yields the orientations of principal curvature. We will adopt the convention $\kappa_1 \geqslant \kappa_2$. (Notice that this definition of principal curvatures is simpler than the Euclidean equivalent, it is the proper definition for the group of motions of the singly isotropic Cayley–Klein space.)

Now we define (Koenderink and van Doorn, 1992) the Casorati curvature C (Casorati, 1889) and the shape index S (Koenderink and van Doorn, 1992) as

$$C = \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}} = \sqrt{r^2 + s^2 + t^2}, \tag{22}$$

$$S = \arctan\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2} = \arctan\frac{t}{\sqrt{r^2 + s^2}}. \tag{23}$$

In terms of the original coefficients a_{ij} one has

$$C = \sqrt{\frac{1}{2}\left(a_{20}^2 + 2a_{11}^2 + a_{02}^2\right)}, \tag{24}$$

$$S = \arctan\frac{a_{20} + a_{02}}{\sqrt{(a_{20} - a_{02})^2 + 4a_{11}^2}}. \tag{25}$$

In computer vision the Casorati curvature occurs as the ‘bending energy’ or ‘total curvature’, it is obviously a differential invariant. The shape index does not depend on size, thus measures pure shape. It is invariant against arbitrary similarities.

A crucial observation derives from the identity

$$\begin{pmatrix} C\sin S \\ C\cos S\cos 2a \\ C\cos S\sin 2a \end{pmatrix} = \begin{pmatrix} t \\ r \\ s \end{pmatrix}. \tag{26}$$

If we use a Cartesian 'shape space' with coordinates $\{r, s, t\}$ then, in polar coordinates, the Casorati curvature is the radius, the shape index the elevation, and double the orientation of the principal directions (no matter which one) the azimuth.

Since the Casorati curvature is a pure size measure, we can use the unit sphere as the space of all shapes and orientations (Fig. 6). The 'north pole' has the convex umbilics, the south pole the concave umbilics, the equator the symmetrical saddles. At the $\pm 45°$ latitudes we have cylinders. A latitude circle contains all orientations of a given shape, whereas a meridian contains all shapes for a given orientation (Fig. 7).

This is the most intuitive and useful representation of quadrics from the perspective of pictorial shape. It is crucial in local shape (i.e., osculating quadric) estimation since it allows a metrical description of shape space (Griffin, 2007), as required in Bayesian inference.

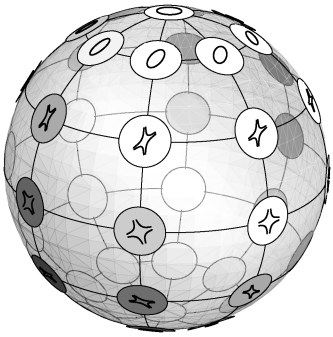

Figure 6. The unit sphere in shape space represents all quadrics of unit Casorati curvature. Here the quadrics are drawn by way of their 'Dupin indicatrix' (Coxeter, 1989), that is the locus $a_{20}x^2 + 2a_{11}xy + a_{02}y^2 = \pm\varepsilon^2$. The poles represent the umbilical (spherical) shapes, the equator the symmetrical saddles, and the 45° latitude circles the cylinders. Along a latitude circle one finds all orientations of a given shape, along a meridian all shapes of a given orientation. Changing the Casorati curvature merely involves changing the diameter of the sphere.

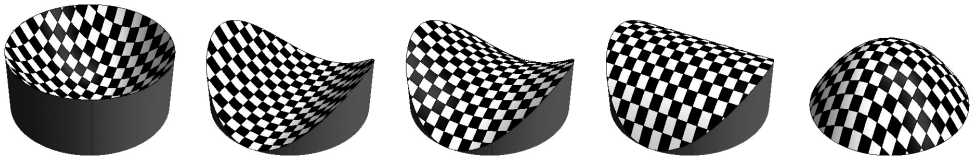

Figure 7. Quadrics along a meridian of the unit sphere in quadrics shape space. From left to right a concave umbilical, a concave cylinder, a symmetrical saddle, a convex cylinder and a convex umbilical.

5.1.2. Local Shape Inference

Assume we do not know the light flow direction and we are interested in shape. (Here we use 'shape' to include orientation, it will rarely be of interest to see a cylinder without seeing its orientation. Such a vision would be conceptual, not optical.) Thus, we are searching for a point on the quadrics shape sphere. Here we can use equation (9). The possible shapes are constrained to lie on a certain plane through the origin. In terms of the $\{r, s, t\}$ parameters the constraint is

$$a_x \Phi_y (r + t) + (a_y \Phi_y - a_x \Phi_x) s + a_y \Phi_x (r - t) = 0. \tag{27}$$

We may as well select coordinates such that the gradient is along the first coordinate direction. Setting $\Phi_x = 1$, $\Phi_y = 0$, we obtain

$$a_y (r - t) - a_x s = 0. \tag{28}$$

Thus, possible inferences lie on a certain great circle of the quadrics shape space. Apparently we have a one-parameter set of possible inferences. This is very useful, for apparently we need only a single additional constraint to settle the case. The ambiguity is not that bad.

From equation (28) it follows that all constraint planes share a common diameter, given by $s = r - t = 0$ or $\{1, 0, 1\}/\sqrt{2}$. Thus, this common diameter lies on the 45° latitude circle, its azimuth being determined by the direction of the gradient. The direction of the light flow then determines which circle out of the bundle one has. The pole of such a circle is given by (by differentiation of equation (28) with respect to r, s and t) $\{\sin\varphi, -\cos\varphi, -\sin\varphi\}/\sqrt{1 + \sin^2\varphi}$, where φ denotes the direction of the light flow (Fig. 8).

This immediately shows that only if the light flow is along the brightness gradient one has umbilicals among the possible inferences, when it is perpendicular to the brightness gradient the only non-hyperbolic (saddle-shaped) solution is a cylinder, whereas saddle-shaped (and cylindrical) inferences are possible in any case.

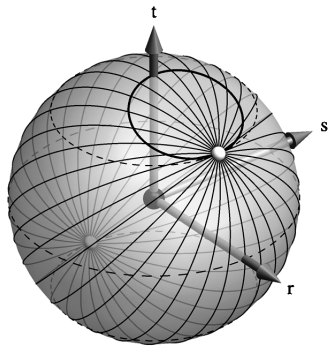

Figure 8. The unit sphere in *rst*-space. An observation of the local contrast gradient due to shading limits possible shape inferences to one of the bundle of great circles sharing the diameter $s = r - t = 0$. The locus of most spherical inferences is the locus drawn fat. The loci of symmetrical saddles (equator) and cylinders (±45° latitude circles) have been drawn dashed. It is very close to a small circle of radius $\pi/8$ and center $\{1, 0, 2\}/\sqrt{5}$.

In the absence of additional information we can only proceed by guessing one parameter (of course the mainstream would perform a Bayesian inference (Bayes, 1763; Brunswik, 1955; Knill and Richards, 1996; Purves *et al.*, 2001), but this is not different from a best guess either). Obvious candidates are the light flow direction and the shape index. If we guess the light flow direction we obtain a unique shape inference, if we guess the shape index, we may infer a light flow direction, but there exists the additional possibility that the guess is an impossible one, ruled out by the optical data.

Suppose you know the light direction. Then the solution must lie on a certain meridian, and you arrive at a unique solution, because the meridian is certain to meet a generic great circle in a single point.

Now suppose you know the shape index. Then the inference must lie on a certain latitude. Thus, you obtain either no solution (the shape index is out of the range of shape indices described by the great circle), or you obtain a pair of solutions. It is a priori clear that umbilics will essentially never be among the set of solutions (only meridians reach the poles), whereas the symmetrical saddles always will (the equator meets any great circle). But even cylinders ($\pm 45°$ latitude) should always be possible, simply take their orientation as different from the gradient direction. There evidently exists a range of possible shape inferences that lies symmetrically about zero (Figs 9 and 10).

Although it will generally not be possible to infer an umbilical if the light flow direction is known, it is a well-defined problem to find the most umbilical inference, that is the one for which the shape index reaches an extremum. Given w_{xx} and w_{xy}, the lacking coefficient w_{yy} is taken to be $w_{yy} = (w_{xx}^2 + 2w_{xy}^2)/w_{xx}$ and the shape is guaranteed to be elliptical. The locus of these inferences on the shape sphere is shown in Fig. 8. It is the intersection with the elliptical cone $\{1 - \cos 2\varphi, -\sin 2\varphi, 2\}$. The locus is almost circular, with (spherical) semi axes $\pi/4 = 45°$ and $\arccos(2/3) \approx 48.19°$.

Figure 8 neatly sums up the possibilities as constrained by a single observation of the local contrast gradient.

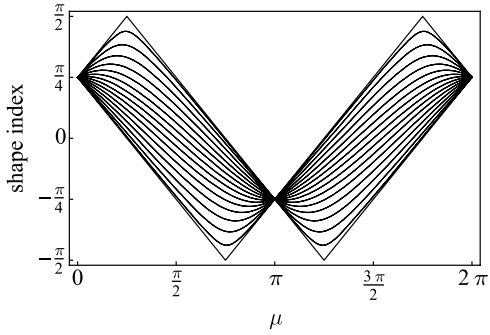

Figure 9. Variation of the shape index inference as a function of the arc length along the constraint circles. The extreme values are shown in Fig. 10.

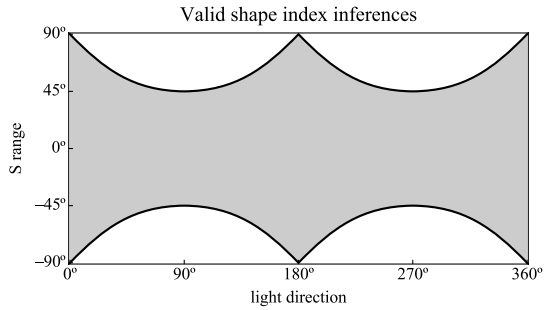

Figure 10. Possible shape index inferences depending upon the light direction relative to the gradient direction. Only if these directions are colinear are umbilical inferences possible. An elliptical shape is almost always possible, at worst one needs to assume a cylinder.

Of course, all this can easily be reformulated in a Bayesian framework if such is desired. A problem is to obtain the priors. In the absence of any knowledge the prior for the flow direction would appear to be the uniform distribution. For biological observers this should probably be skewed towards a preference for illumination from above, the precise prior most likely being idiosyncratic. In the simplest case the probability density could be taken as

$$P(\vartheta) = \frac{1}{2\pi}(1 + a\cos\vartheta), \tag{29}$$

where the parameter a, with $0 < a < 1$, is observer dependent, and the angle ϑ parameterizes the direction of light flow, $\vartheta = 0$ representing illumination from above. A prior for the shape index can be obtained from first principles (Koenderink and van Doorn, 2003b; Lillholm and Griffin, 2009), for instance one might use the distribution of shape indices for isotropic random Gaussian surfaces. One finds

$$P(S) = \frac{2\sqrt{2}\cos S}{(3 + \cos 2S)^{3/2}}. \tag{30}$$

Such a distribution peaks at the cylinders ($S = \pm\frac{\pi}{4}$) and is zero for the umbilicals ($S = \pm\frac{\pi}{2}$). The probability to encounter a saddle shape is $1/\sqrt{3} \approx 57.7\%$, to encounter a convexity (same for concavity) 21.1%. We do not use the Bayesian framework here because it implicitly assumes the third person account, thus, it is not consistent with a view from the inside.

5.2. Global Inferences

Global inferences are possible if you know the light flow direction. In that case you pick coordinates such that $a_x = 1$, $a_y = 0$, and you obtain w_{xx} and w_{xy} up to a common factor. The common factor is irrelevant, since it involves only the depth of relief. Thus, you may limit the absolute magnitude or force $w_{xx}^2 + w_{xy}^2 = 1$, whatsoever the preference, it does not matter. In any case, you obtain w_{xx} and w_{xy}.

Figure 11. Ribbon geometry. The ribbon at left is planar and only modulated through geodesic curvature, the ribbon at center is only modulated through normal curvature, and the ribbon at right is purely twisted. A generic ribbon will have geodesic and normal curvature, as well as twist. Ribbons assume an ontological position between curves and surfaces.

Notice the ambiguity implied by the necessary integration. If $w(x, y)$ is a solution, then so is $w'(x, y) = w(x, y) + Cx + F(y)$, where C is an arbitrary constant, and $F(y)$ an arbitrary function of y.

One way to proceed towards a solution is to notice that the observation concerns the gradient of w_x, for $\nabla w_x = \{w_{xx}, w_{xy}\}$. Integrating the gradient is a well-defined problem, and integrating the result once again over the x-direction yields the relief, up to an arbitrary 'additive plane' and an arbitrary depth scaling. These ambiguities are understood anyway, thus, this is a solution. The integration of the gradient can be formulated as a simple least squares problem, a numerical procedure would simply use the singular values decomposition. It is perhaps a method that is biologically less likely though.

Another method is more geometrical and might be implemented in a biologically more plausible manner. You do a one-dimensional integration (integrating the gradient is a two-dimensional integration) along the field lines of the light flow. You interpret w_{xx} as the normal curvature of a strip, and w_{xy} as the twist of the strip (Fig. 11). The strips are of course again subject to arbitrary additive planes and depth scalings. Suppose you integrate only w_{xx}, you obtain

$$\iint w_{xx}(x, y)\,\mathrm{d}x\,\mathrm{d}x = C_0(y) + C_1(y)x + w(x, y), \tag{31}$$

where $C_{0,1}$ are arbitrary integration constants that still depend upon y. If you integrate over the area of a flow box (Fig. 12), you may assume an arbitrary depth along the boundary of the box, this settles the constants of integration.

This shows the importance of information concerning the boundary of the flow box (Fig. 12). It is crucial in typical empirical results reported in the literature. The most popular stimuli are linear gradients in a circular area, embedded in a uniform surround. The microgenetic process will almost certainly interpret the circular boundary as a planar edge at which the relief meets the flat (because it has uniform brightness) environment. Then the umbilical solution is forced, and indeed, observers report only convex and concave umbilicals (Palmer, 1999; Ramachandran, 1988). It need not surprise that a change of shape of the boundary will affect the resulting relief. A square boundary lined up with the flow yields an apparent cylinder, for instance. Such results are evidently due to the boundary conditions, they have only minor relevance to the shading cue proper.

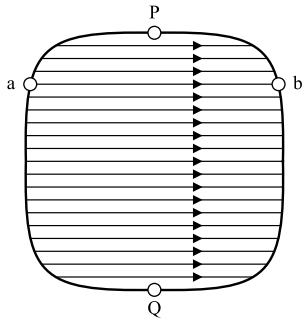

Figure 12. A 'flow box'. At all points except P and Q the boundary is transverse to the flow. By specifying arbitrary depths along the contour one uses the full room of ambiguities. Consider the flow line ab, you can both add a constant and a tilt in the flow direction by assigning suitable depth values to the boundary points a and b.

6. Some Consequences of the Theory

In this section we demonstrate the formal theory through application on a number of commonly presented stimuli. We also show some results for photographs of real scenes (Figs 13 and 14).

6.1. Estimation of Flow, Splay and Swerve

In Fig. 13 we show a photograph of a concrete sphere. The object is illuminated by the sun, the edge of the body shadow is clearly visible, and, because we have the spherical shape as ground truth, is sufficient to estimate the elevation of the source as 28°. The concrete yields plenty of textural detail, thus running a structure tensor analysis we obtain a reasonably good estimate of the light flow.

The photograph measures 512×512 pixels. The spatial derivatives for the structure tensor were computed at a one-pixel scale, whereas the local averaging was at a 16-pixel scale. Such operations are easily available in the human visual cortex (Hubel, 1989; Koenderink and van Doorn, 1990; Tolhurst, 1972), thus, a flow field of this nature may be assumed available to the microgenetic process. The field is close to that theoretically expected, except on the lower half of the sphere, where deviations are likely to be due to scattered radiation from the environment. In any case, the flow direction at the center is easily estimated with an accuracy of a few degrees.

The flow field is mainly splay with only a little swerve. At the center we estimate the splay to be about half a radian (ca. 30°) turn for a sphere radius. (A more precise analysis yields 0.47 radians per radius, not essentially different from the rough estimate by eye.) The swerve is at least an order of magnitude smaller.

6.2. Purely Local Inference

From the flow field shown in Fig. 13 we immediately conclude that the surface is convexly curved in the direction orthogonal to the flow, whereas the mixed second-order derivative is essentially zero. A numerical estimate yields a radius of curvature

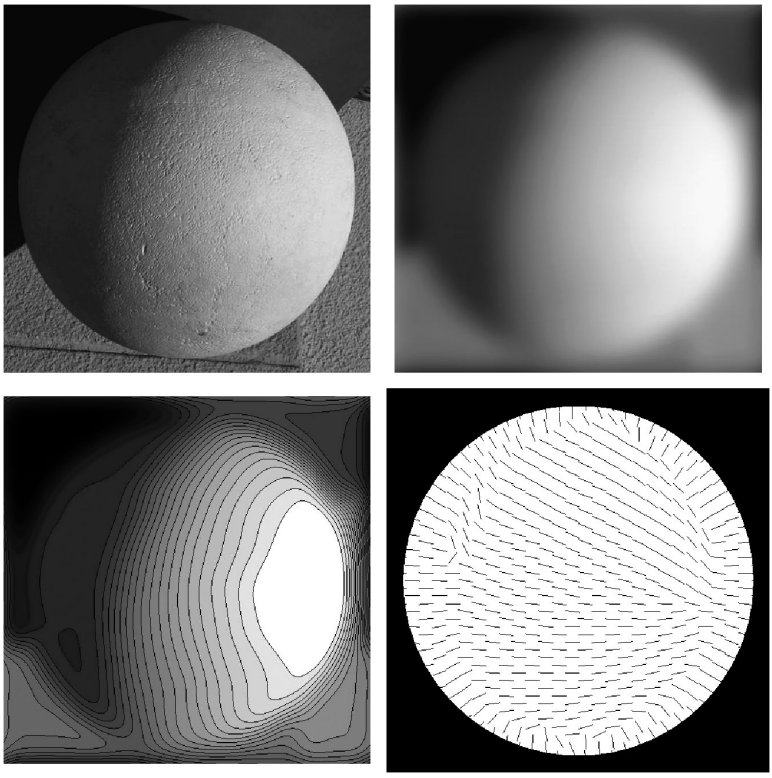

Figure 13. A concrete sphere, illuminated by sunlight, photographed along a San Francisco street. (Upper left) The cut-out from the photograph (rotated so as to put the flow direction roughly horizontal); (upper right) a blurred version; (lower left) the isophotes of the blurred version; (lower right) the estimated flow.

of approximately 1.12, only a little higher than the true value (exactly one). This is a purely local estimate (no integration), purely on the basis of flow.

Next consider the shading proper. From the isophote pattern of the blurred photograph the contrast gradient is seen to be lined up with the flow direction. Thus, there is significant curvature along the flow, whereas the second-order mixed derivative is negligible. A numerical estimate yields a contrast gradient of 0.50, which, combined with the elevation of the source, yields an estimate of the radius of curvature of 1.05, only little higher than the true value (exactly one). This, again, is a purely local estimate (no integration), this time purely on the basis of shading.

If the microgenetic process does not use the elevation of the source (which is available only through the edge of the body shadow, the boundary of the sphere, and the assumption of sphericity), it will still classify the surface as convex umbilical at the center (combining the splay and shading cues). Combined with the circular outline (interpreted as occluding contour) this is all the microgenetic process needs to let the observer hallucinate a sphere. This is a presentation that does not conflict

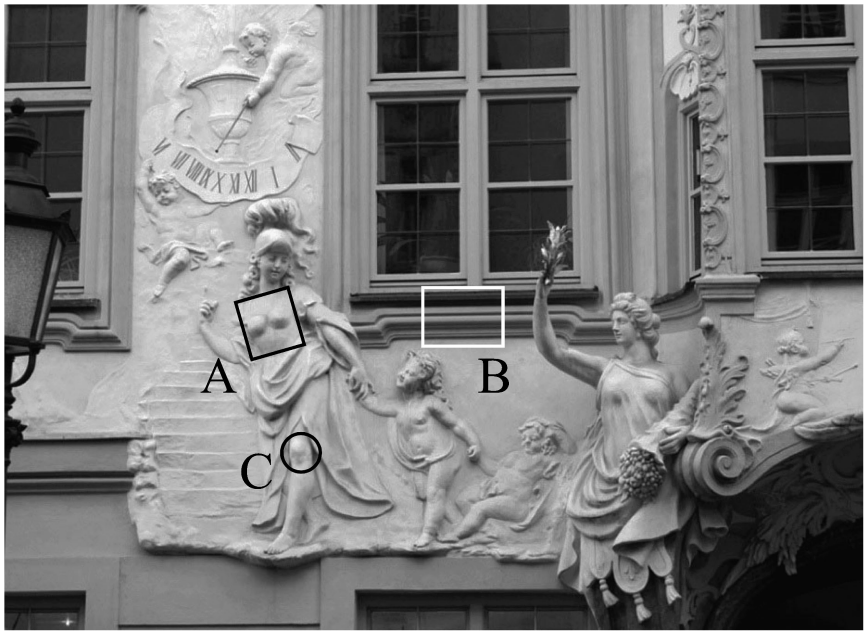

Figure 14. Facade of the Asam house at Munich, a snapshot taken from the other side of the street. Illumination by the overcast sky, due to vignetting by the facades at both sides of the street mainly from above. The ornamentation is in white stucco, some in low relief, at other places fully modeled in the round. Indicated regions A, B and C are considered in the text.

with any of the optical structure, and thus makes sense. Moreover, it is useful as it identifies sphericity. Combined with the contour it settles the scene.

Notice that these inferences are very robust. The concrete is not Lambertian, the texture and city dirt perturb the shading distribution (as evident from the shapes of the isophotes, etc.). It is unlikely that a full fledged CV analysis of the photograph would do much better.

In Fig. 15 we show area C from Fig. 14 and a local shape inference based on shading. In this case the flow is not available from texture, we used the 'light from above' assumption, which agrees with the overall structure of the photograph. With shading alone one has to guess at the shape index, here we used the 'most spherical' inference. It turns out to be almost umbilical, not surprising given the isophotes pattern. This is about the best the microgenetic process can do with scarce optical structure. Even so, the inference is clearly a useful one.

6.3. Inferences Along Flow Lines

Integrating along flow lines is useful in general, but perhaps especially so in cases of cylindrical relief. In the latter case one needs not even estimate the flow direction since any generic flow direction (transverse to the cylinder axis) will yield the same result up to sign. We illustrate this case on area B of Fig. 14.

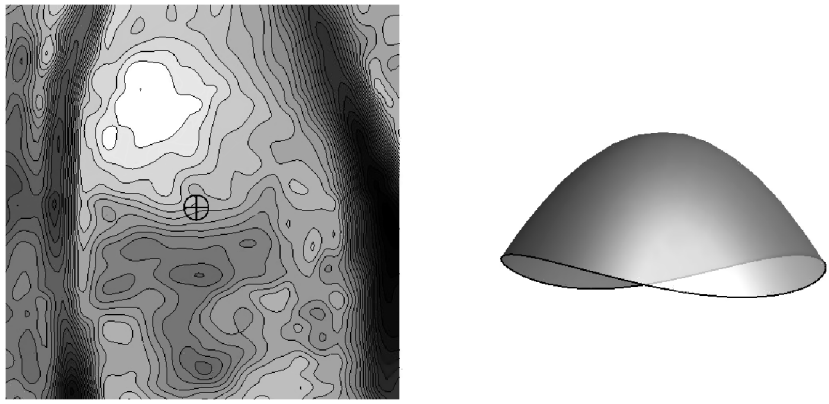

Figure 15. Area C from Fig. 14 (the exposed knee of the standing figure) and a local shape inference (at the point marked at the center of the image) based on shading.

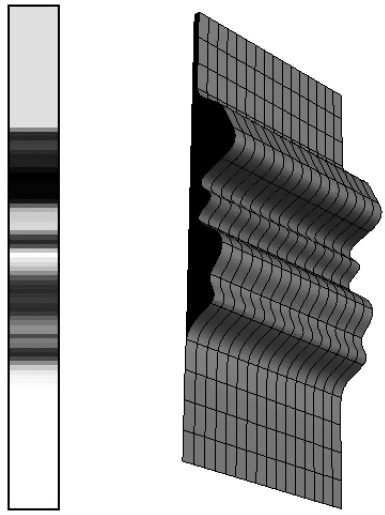

Figure 16. Area B from Fig. 14 (the moulding of the lower window sill) and a local shape inference based on the integration of the contrast gradient along the vertical.

We simply integrate the contrast gradient along the vertical and obtain the profile shown in Fig. 16. There is no way to compare the result with physical reality, but — for this paper more importantly — it looks like what one would guesstimate by eye measure. Of course the result is only defined up to the usual ambiguities. In this case the vertical plane of the facade lifts all ambiguities except for the depth of relief.

Such simple inference engines are easily implemented in biological systems, and are surprisingly useful. Many shapes of immediate biological interest are approximately cylindrical, think of arms, legs, sections of faces or thoraxes, treetrunks and so forth (Marr, 1982). In such cases it is not even necessary to estimate the flow,

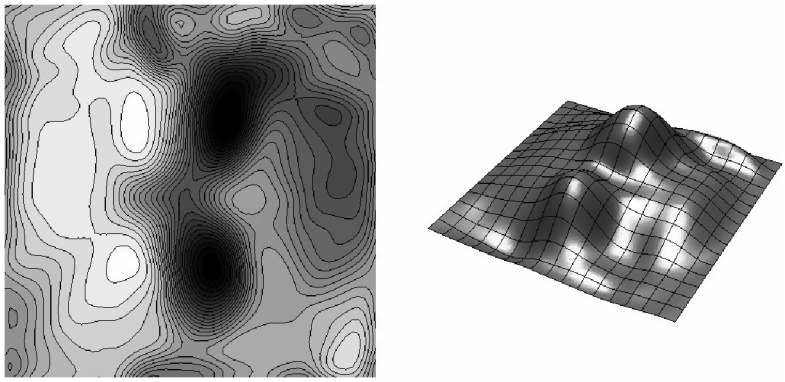

Figure 17. Area A from Fig. 14 (the upper part of the thorax of the standing figure) and a global shape inference based on the integration of the contrast gradient over the rectangular area.

one simply integrates the contrast gradient along any direction that is transverse to the (approximate) cylinder axis.

6.4. General (Extended, but Non-Cylindrical) Reliefs

The area A from Fig. 14 has been selected because it is evidently structured in two dimensions, and can in no way be approximated with a cylindrical moulding without losing crucial structure.

We defined a rectangular area, tilted such that one side is (by eye measure) roughly parallel to the light flow. Then we integrated once to obtain the depth gradient, and another time to obtain the depth. The result is shown in Fig. 17. It clearly reflects the shape impression of the upper part of the torso of the standing figure quite well. Of course, the result is subject to the usual ambiguity transformations, but the default (frontoparallel) posture works well enough in this case. One could no doubt do better by segmenting out the torso and providing suitable boundary conditions. However, the quick and dirty method is quite effective, and probably the microgenetic process required in order to present the observer with an effective hallucination of the depth dimension.

6.5. The Importance of Boundary Constraints

The conventional stimulus for shape from shading psychophysics is a circular disk filled with a linear brightness gradient. Common wisdom has that it can be perceived either as a convex or a concave umbilical quadric (either a ‘sphere’ or a spherical ‘cup’) (Palmer, 1999; Ramachandran, 1988). It has been noticed that the shape of the boundary makes a difference, in that a square lined up with the gradient looks like a cylindrical surface (sometimes referred to as ‘illusory’), and an equilateral triangle lined up with a median line like a conical surface. It is not generally recognized that SFS algorithms yield many more possible interpretations, even for the conventional circular disk. This is also the case for the formalism presented

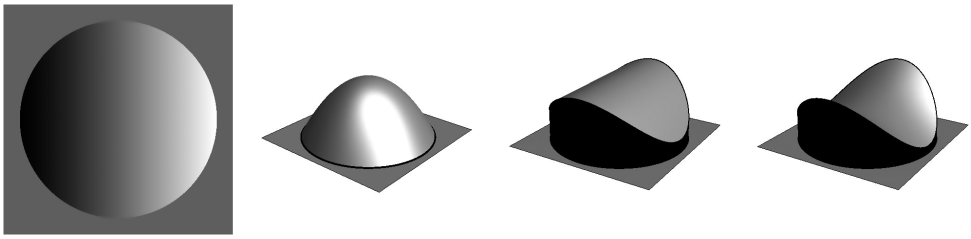

Figure 18. The conventional SFS stimulus and three equally possible inferences (an umbilical, a cylinder and a symmetric saddle, infinitely more are possible). Only the umbilical shape is reported.

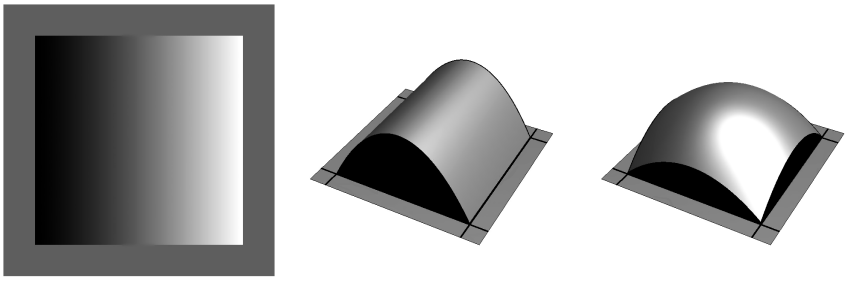

Figure 19. The same gradient as that in Fig. 18, but presented in a square outline, lined up with the gradient. The cylindrical inference is preferred over the umbilical one, apparently because of the simplicity of its outline in depth. Notice that part of the boundary of the flow box is parallel to the flow, such parts play no role in the inference.

here as a model for SFS as seen from the inside. In Fig. 18 we show a few of the (infinite) possibilities.

It seems likely that the microgenetic process applies rather strong boundary conditions (Minsky, 1974). For instance, the assumption that the outline of the disk is a planar, frontoparallel curve in pictorial space rules out all interpretations but the umbilical surfaces. But this would imply that perceptions on the basis of such stimuli reveal more about shape from contour than about shape from shading. Thus, this is a conceptually very important issue.

As discussed earlier, the depth structure of the boundary of a flow box can be prescribed arbitrarily. This immediately allows us to explain the perception of cylinders and cones in rectangular or triangular outlines filled with the same linear gradient as the conventional circle (Figs 19 and 20). The resulting presentations are perfectly valid interpretations, just as the umbilical interpretation is. For non-circular outlines it may be the case that non-spherical inferences are in agreement with particularly simple depth structures of the outline (e.g., a planar curve), whereas an umbilical inference might imply a much more complicated structure.

7. Shading in the Visual Arts

Shading in the visual arts (Hamm, 1982; da Vinci, before 1542) is quite distinct from shading in computer graphics (CG) (Foley *et al.*, 1990). Of course, the ob-

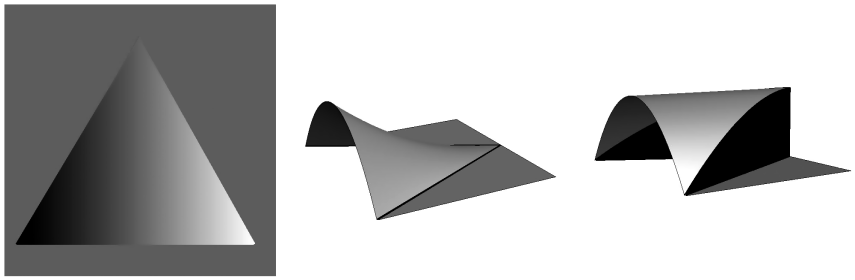

Figure 20. The same gradient as that in Fig. 18, but presented in a triangular outline, lined up with the gradient. Two possible inferences are shown. The one at the center has a flat outline, but is hyperbolically curved. Observers tend to report 'seeing a cone', but quantitative data as to the perceived shape are not available. None of the solutions is a true right circular cone.

jectives are also different. Mainstream CG aims at veridical, or photographic, renderings of virtual scenes. Various deviations from veridicality are accepted due to considerations of computational overhead, but as computing power grows, methods are invariably moving in the direction of increased veridicality.

The visual artist typically has different objectives. Although some make great efforts at photographic (or preferably better!) effects, most are concerned with the picture as a vehicle of evoking certain intended responses in the audience. This rarely involves the issue of veridicality or only marginally so. Important objectives are the tonal nature of the picture and the composition. We are mainly concerned with the shading of virtual (imagined) scenes here, 'optical' painting *alla prima* is quite another issue. However, true optical painting is rare. Even in front of a model most artists will shade according to their usual 'system', rather than acting like a camera. Most contemporary artists consider the camera as a tool as a matter of course (and so does their audience), so one feels rarely an incentive to act like one.

Typical shading methods will not directly use Lambert's cosine law (Lambert, 1760) at all. Few artists would be able to define the cosine function and its relation to Euclidean geometry. Those that do tend to have other interests besides the art. There are two major methods (each with many variations) in common use.

One method sees surfaces as (variously modulated) ribbons (Bridgman and Simon, 2001; Hamm, 1982) (Fig. 21(left)). The ribbons are only conceptual, any surface can be analyzed into ribbons in many ways. One picks a ribbon that runs along the (often imagined) direction of illumination. As the ribbon turns away from the light one darkens it, if it turns towards the light one lightens it. This is a very effective method. It yields a first handle on the shading that will be modulated or corrected, in numerous minor ways, but the initial impact of the picture is due to the dominance of the ribbon and the artist is careful not to destroy that. The observer should not notice the method, but 'feel' the ribbon structure in subsidiary awareness.

The other method analyzes complicated surfaces or shapes in terms of 'ovoids', that is convex, egg-shaped bodies (Hatton, 1904) (Fig. 21(right)). The actual sur-

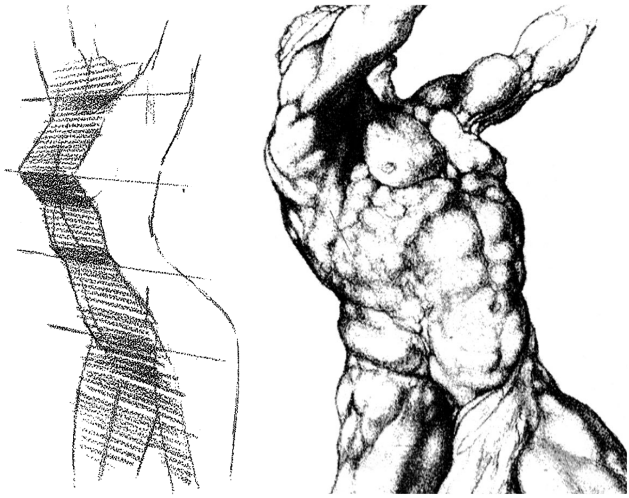

Figure 21. (Left) Example of the method of ribbons (Hamm, 1982). Here the ribbon has been chosen to run vertically downwards over the front of the torso, but other choices are equally possible. When artfully blended, the ribbon trick becomes invisible to the lay observer. The method by nature 'generalizes' the shape, only major landmarks can be identified and named. (Right) Example of the method of ovoids. Rimmer (Rimmer, 1970) (the artist) was an expert on the muscular anatomy. The ovoids (main contractile bellies) have been artfully blended (but Rimmer could name each and any of them). Notice how the torso looks much like a sack of potatoes.

face will of course be suitably smoothed, but the basic ovoids remain visible (the observer is supposed to 'feel' them in subsidiary awareness), a bit like a sack of potatoes seen from the outside. Since shading of an egg is trivial, this is a great method to generate the typically far more complicated shading of an actual surface.

Here 'darker' and 'lighter' are relative to the tonal scale of the painting, they have little to do with the actual illuminances or radiances.

The shading account given here fits in very well with such practices. It can easily be implemented in CG style, and yields a 'shading method' that deviates from Lambert's cosine law (Pottmann, Hagen and Divivier, 1991). Here is a summary account of such a method. Consider a surface $z(x, y)$. The distance gradient is $\mathbf{g} = \{z_x(x, y), z_y(x, y)\}$. Pick a light flow direction (in the picture plane!) $\mathbf{i} = \{i_x, i_y\}$ and define the brightness as $\mathbf{g} \cdot \mathbf{i}$. Notice how this is different from the 'real thing': the light direction lives in the picture plane instead of space, and so does the attitude of the surface. The brightness is a number that ranges over the real line, $\{-\infty, +\infty\}$. The advantages are that one needs not worry about body and cast shadows at all. In order to draw the picture (on the screen) you need to map the real line on the range of gray-tones that the machinery will display. A convenient way is to define

$$g = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{b-m}{r}\right)\right) \tag{32}$$

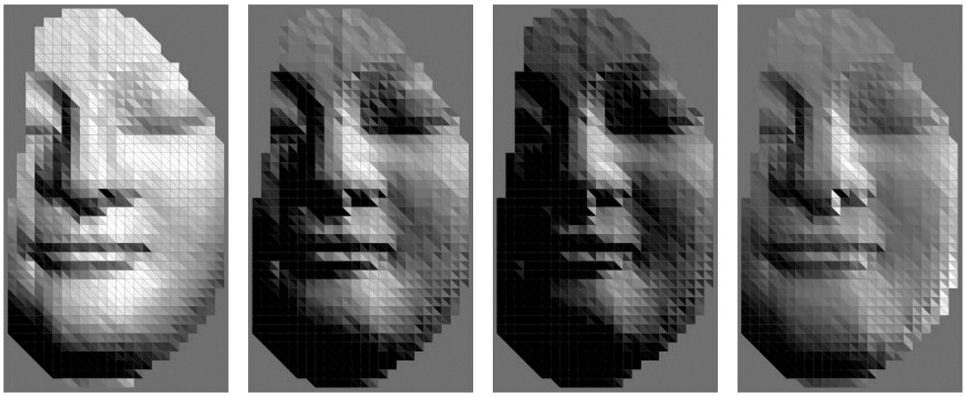

Figure 22. The sculpture has been crudely triangulated (due to the late John Willats (Dubery and Willats, 1972)), each (of course, planar) triangle is illuminated according to Lamberts cosine law in the three leftmost cases, the light vector moving from almost frontal to almost sideways. In the latter cases many triangles lie in the body shadow region and are rendered flat black. The image on the right has been rendered *via* the gradient method. In this method body shadows have no meaning. Although not based on physics and Euclidean geometry, such renderings look perfectly acceptable to the human observer.

(erf(x) the conventional error function) where you can assign the mid-point m and range r of rendered brightness b so as to arrive at gray-tones g in the range (0, 1) (Fig. 22). The parameterization gives desirable freedom in the tonal composition of the picture. If so desired there are obvious possibilities to skew the gray-tone distribution, and gain even more control.

Such possible CG methods are close to the shading account discussed here and are much closer to artistic praxis than the 'correct' methods. In scientific visualization they also make much more sense than the Euclidean methods (Pottmann, Hagen and Divivier, 1991) because there usually is no natural Euclidean setting. The case of human perception is similar. 'Depth' is not a Euclidean dimension, so pictorial space cannot be approached by way of the 'correct' framework at all. For the view from the inside, the Euclidean method is quite alien, involving many unwarranted assumptions.

8. Relation to the Formal Methods of CV

In order to show up the formal relation we use the simple example of a one-dimensional image. The general case yields no additional insights and is troubled by much longer expressions. Consider a surface

$$z(x) = a + \tan\vartheta\, x + \varepsilon\zeta(x), \tag{33}$$

where $\zeta(0) = 0$, $\zeta_x(0) = 0$ (these conditions are not essential, but formally very convenient). This describes the surface by absolute distance a, slope ϑ and some additional modulation described by $\zeta(x)$. The parameter ε is introduced to keep

track of orders of magnitude, for the moment one may simply take $\varepsilon = 1$, eventually we consider the limit $\varepsilon \to 0$. Consider the light direction to be given by

$$\mathbf{i} = \cos(\vartheta + \varphi)\mathbf{e_x} + \sin(\vartheta + \varphi)\mathbf{e_z}, \tag{34}$$

where the parameter φ describes the elevation of the source relative to the overall slope of the surface. For the moment consider $\varphi > 0$, eventually we will consider the limit $\varphi \to 0$.

The direction of the surface normal is $\{-z_x(x), 1\}$, thus one has

$$\mathbf{n}(x) = \frac{\{-\tan\vartheta - \varepsilon\zeta_x(x), 1\}}{\sqrt{1 + (\tan\vartheta + \varepsilon\zeta_x(x))^2}}, \tag{35}$$

and finds the illuminance, invoking Lambert's cosine law, as $I(x) = \mathbf{i} \cdot \mathbf{n}(x)$. The illuminance contrast is defined as

$$c(x) = \frac{I(x) - I(0)}{I(0)}, \tag{36}$$

such that $c(0) = 0$. The expression of interest is the gradient of the illuminance contrast

$$\nabla c(0) = -\varepsilon \cos^2\vartheta \cot(\varphi)\zeta_{xx}(0). \tag{37}$$

Now consider the limit for low relief and almost raking illumination. It is a double limit in which we want to keep the ratio ε/φ at a finite value, whereas both ε and φ approach zero. Setting the ratio to some constant and taking the limit one obtains

$$\nabla c(0) \propto \zeta_{xx}(0). \tag{38}$$

This is exactly the result from the previous calculation (equation (4)). Thus, the Euclidean geometry with Lambert's cosine law contains the previous formalism as the limit for shallow relief and raking illumination.

It is a very useful approximation for a number of reasons. For instance, the slope of the surface does not matter, one has no perspective foreshortening issues. There is no need to worry about body and cast shadows. Various calibration issues play no role. This renders the method attractive for biological purposes, the more so because everything can be framed in terms of the brightness and pictorial spaces. The latter fact need not surprise one, it is no doubt one of the reasons why these spaces have the structure they do.

9. Conclusions

We have presented a speculative account of the shading cue as it appears in the first person account, a 'view from the inside'. The general theory implies a number of ways to handle shading patterns, depending upon the availability of additional structure (such as light flow, shape or spatial attitude), and situational awareness. This makes sense, because in no case known to us is a perceived shape causally dependent upon the shading cue alone. In fact, it may well be impossible to design a

stimulus that would really isolate the shading cue (Erens, Kappers and Koenderink, 1993). In virtually all cases the stimulus bouquet, including shading, will severely underdetermine possible inferences. Since perception (whether of the human observer or the autonomous robot of AI) does not have the option to quit, one needs to consider ways to proceed in such cases.

The theory allows one to develop many rather immediate predictions. For almost any given case there is an obvious implementation of the theory. Usually there will be a number of options left open. Whether an option is used and how is a matter of hypothesis, and the theory will allow one to put it to the test. An example is the case of the linear gradient inside a circular, rectangular or triangular boundary. Here the human observer apparently goes for the most simple (planar) outline in depth, simply ignoring the infinitely many equally valid inferences.

Whether it is desirable to model the inner perspective of an autonomous robot on the human condition is a moot point. Some streams in AI have the explicit aim to mimic human capabilities, but there exists a broad realm of enquiry here. The present formalism is general enough to offer a useful platform to start from.

References

Adams, A. (1950). *The Print: Contact Printing and Enlarging*. Morgan and Morgan, New York.

Adelson, E. H. and Bergen, J. R. (1991). The plenoptic function and the elements of early vision, in: *Computational Models of Visual Processing*, M. Landy and J. A. Movshon (Eds), pp. 3–20. MIT Press, Cambridge, MA.

Albertazzi, L. (2008). The ontology in perception, in: *Theory and Applications of Ontology (TAO)*, R. Poli, J. Seibt and J. Symons (Eds), Vol. 1. Springer, Berlin.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions, Giving Some Account of the Present Undertakings, Studies and Labours of the Ingenious in Many Considerable Parts of the World* **53**, 370–418 (by the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S.).

Belhumeur, P. N., Kriegman, D. and Yuille, A. (1999). The bas-relief ambiguity, *Int. J. Comp. Vis.* **35**, 33–44.

Berkeley, G. (1709). *An Essay Towards a New Theory of Vision*. J. Pepyat, Dublin.

Block, N. (2003). Mental paint, in: *Reflections and Replies, Essays on the Philosophy of Tyler Burge*, M. Hahn and B. Ramberg (Eds). MIT Press, Cambridge, MA.

Born, M. and Wolf, E. (1999). *Principles of Optics*. Cambridge Univ. Press, Cambridge, UK.

Brentano, F. (1995). *Psychology from Empirical Standpoint*. Routledge, London (original 1874).

Bridgman, G. B. and Simon, H. (2001). *Bridgmans Complete Guide to Drawing from Life*. Sterling Publishing, New York.

Brown, J. W. (2000). *Self-Embodying Mind: Process, Brain Dynamics and the Conscious Present*. Midpoint Trade Books, Barrytown, NY.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology, *Psychol. Rev.* **62**, 193–217.

Casorati, F. (1889). Nuova denizione della curvature della supercie e suo confronto con quella di Gauss, *Rend. Inst. Matem. Accad. Lomb.* **2**, 335–346.

Cayley, A. (1859). Sixth memoir upon the quantics, *Philos. Trans. Roy. Soc. Lond.* **149**, 61–70.

Chantler, M. J., Russell, G. T. and Linnett, L. M. (1994). Illumination: a directional filter of texture?, *Br. Mach. Vis. Conf.* **2**, 449–458.

Coxeter, H. S. M. (1989). *Introduction to Geometry*. Wiley, New York.

da Vinci, L. (before 1542). Codex Urbinas, see http://en.wikipedia.org/wiki/Codex_Urbinas.

Dubery, F. and Willats, J. (1972). *Drawing Systems*. Publisher Studio Vista Books, London.

Erens, R. G. F., Kappers, A. M. L. and Koenderink, J. J. (1993). Perception of local shape from shading, *Percept. Psychophys.* **54**, 145–156.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf und Härtel, Leipzig.

Foley, J. D., van Dam, A., Feiner, S. K. and Hughes, J. F. (1990). *Computer Graphics: Principles and Practice*. Addison-Wesley, Reading, MA.

Forsyth, D. A. and Ponce, J. (2002). *Computer Vision: A Modern Approach*. Prentice-Hall, Upper Saddle River, NJ.

Gershun, A. (1936). *The Light Field*. Moscow; translated by P. Moon and G. Timoshenko (1939), *J. Math. Phys.* **XVIII**, 51–151.

Griffin, L. D. (2007). The 2nd order local-image-structure solid, *IEEE Trans. Pattern Anal. Mach. Intell.* **29** 1355–1366.

Hamm, J. (1982). *Drawing the Head and Figure*. Perigee Trade, New York.

Hatton, R. G. (1904). *Figure Drawing*. Chapman and Hall, London.

Hildebrand, A. (1893). *Das Problem der Form in der bildenden Kunst*. Heitz and Mündel, Strasbourg.

Hoffman, D. (2008). Sensory experiences as cryptic symbols of a multi-modal user interface, in: *Kunst und Kognition*, M. Bauer, F. Liptay and S. Marschall (Eds), pp. 261–279. Wilhelm Fink, Munich.

Horn, B. K. P. (1970). Shape from shading: a method for obtaining the shape of a smooth opaque object from one view, *PhD Thesis*, MIT.

Hubel, D. (1989). *Eye, Brain and Vision*. W. H. Freeman, New York.

Klein, F. (1871). Über die sogenannte nicht-Euklidische Geometrie, *Math. Ann.* **6**, 112–145.

Klein, F. (1893). Vergleichende Betrachtungen über neuere geometrische Forschungen, *Math. Ann.* **43**, 63–100.

Klein, F. (1928). *Vorlesungen über nicht-Euklidische Geometrie*. Springer, Berlin.

Knill, D. C. and Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge Univ. Press, Cambridge, UK.

Koenderink, J. J. and van Doorn, A. J. (1990). Receptive-field families, *Biol. Cybernet.* **63**, 291–297.

Koenderink, J. J. and van Doorn, A. J. (1992). Surface shape and curvature scales, *Image Vis. Comput.* **10**, 557–564.

Koenderink, J. J. and van Doorn, A. J. (2002). Image processing done right, in: *Proc. European Conference on Computer Vision*, pp. 158–172. Springer, London.

Koenderink, J. J. and van Doorn, A. J. (2003a). Pictorial space, in: *Looking into Pictures: An Interdisciplinary Approach to Pictorial Space*, H. Hecht, R. Schwartz and M. Atherton (Eds), Chapter 12. MIT Press, Cambridge, MA.

Koenderink, J. J. and van Doorn, A. J. (2003b). Local structure of Gaussian texture, *IEICE Trans. Inform. Syst.* **86**, 1165–1171.

Koenderink, J. J. and van Doorn, A. J. (2008). The structure of visual spaces, *J. Math. Imaging Vis.* **31**, 171–187.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L. and Todd, J. T. (2001). Ambiguity and the "mental eye" in pictorial relief, *Perception* **30**, 431–448.

Koenderink, J. J., van Doorn, A. J. and Pont, S. C. (2004). Light direction from shad(ow)ed random Gaussian surfaces, *Perception* **33**, 1405–1420.

Koenderink, J. J., van Doorn, A. J. and Todd, J. T. (2009). Wide distribution of external local sign in the normal population, *Psychol. Res.* **73**, 14–22.

Koenderink, J. J. and Pont, S. C. (2003). Irradiation direction from texture, *J. Optic. Soc. Amer. A* **20**, 1875–1882.

Lambert, J. H. (1760). *Photometria, sive, De mensura et gradibus luminis, colorum et umbræ*. V. E. Klett, Augustæ Vindelicorum.

Lillholm, M. and Griffin, L. D. (2009). Statistics and category systems for the shape index descriptor of local 2nd order natural image structure, *Image Vis. Comput.* **27**, 771–781.

Lorenz, K. (1977). *Die Rückseite des Spiegels*. DTV, München.

Marr, D. (1982). *Vision*. Freeman, New York.

Metzger, W. (1953). *Gesetze des Sehens*. Kramer, Frankfurt am Main.

Minsky, M. (1974). *A Framework for Representing Knowledge*. MIT AI Laboratory Memo 306, Cambridge, MA.

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA.

Pentland, A. (1988). Shape information from shading: a theory about human perception, in: *Proc. International Conference on Computer Vision*, pp. 404–413. IEEE Computer Society Press, Los Alamitos, CA.

Poggio, T. (1984). Low-level vision as inverse optics, in: *Proc. Symposium on Computational Models of Hearing and Vision*, M. Rauk (Ed.), pp. 123–127. Academy of Sciences of the Estonian S.S.R., Tallinn.

Poli, R. (2001). The basic problem of the theory of levels of reality, *Axiomathes* **12**, 261–283.

Pottmann, H., Hagen, H. and Divivier, A. (1991). Visualizing functions on a surface, *J. Visualizat. Comp. Animat.* **2**, 52–58.

Purves, D., Lotto, R. B., Williams, S. M., Nundy, S. and Yang, Z. (2001). Why we see things the way we do: evidence for a wholly empirical strategy of vision, *Philos. Trans. Roy. Soc. Lond. B* **356**, 285–297.

Ramachandran, V. S. (1988). Perception of shape from shading, *Nature* **331**, 163–166.

Riedl, R. (1975). *Die Ordnung des Lebendigen. Systembedingungen der Evolution*. Paul Parey, Hamburg.

Rimmer, W. (1970). *Art Anatomy*. Dover Publications, New York.

Rosenthal, V. (2004). Microgenesis, immediate experience and visual processes in reading, in: *Seeing, Thinking and Knowing*, A. Carsetti (Ed.), pp. 221–245. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Sachs, H. (1987). *Ebene isotrope Geometrie*. Friedrich Vieweg & Sohn, Braunschweig.

Sachs, H. (1990). *Isotrope Geometrie des Raumes*. Friedrich Vieweg & Sohn, Braunschweig.

Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge Univ. Press, Cambridge, MA.

Steenrod, N. (1951). *The Topology of Fiber Bundles*. Princeton Univ. Press, Princeton, NJ.

Strubecker, K. (1941). Differentialgeometrie des isotropen Raumes I, *Sitzungsberichte der Akademie der Wissenschaften Wien* **150**, 1–43.

Strubecker, K. (1942). Differentialgeometrie des isotropen Raumes II, *Math. Zeits.* **47**, 743–777.

Strubecker, K. (1943). Differentialgeometrie des isotropen Raumes III, *Math. Zeits.* **48**, 369–427.

Strubecker, K. (1945). Differentialgeometrie des isotropen Raumes IV, *Math. Zeits.* **50**, 1–92.

Swift, J. (1906). *Gullivers Travels and Other Works*. Routledgen, London.

Tinbergen, N. (1951). *The Study of Instinct*. Clarendon Press, Oxford.

Tinbergen, N. (1975). *The Animal and Its World*, Vol. 1. Harvard Univ. Press, Cambridge, MA.

Tolhurst, D. J. (1972). On the possible existance of edge detector neurones in the human visual system, *Vision Res.* **12**, 797–804.

von Helmholtz, H. (1860). *Handbuch der Physiologischen Optik*. Voss, Leipzig.

von Uexküll, J. (1921). *Umwelt und Innenwelt der Tiere*. Springer, Berlin.

Wohlfahrt, E. (1932). Der Auffassungsvorgang an kleinen Gestalten. Ein Beitrag zur Psychologie des Vorgestalterlebnisses, *Neue Psychol. Stud.* **4**, 347–414.

Yaglom, I. M. (1979). *A Simple Non-Euclidean Geometry and Its Physical Basis*. Springer, New York.

Zhang, R., Tsai, P. S., Cryer, J. E. and Shah, M. (1999). Shape from shading: a survey, *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 690–706.