Chapter 2
Shape, Shading, Brain and Awareness

Jan Koenderink and Andrea van Doorn

Abstract. Shading is one of the generic "monocular depth (and shape) cues". It is of conceptual interest because it apparently implies "causal relations" between the geometry of the scene in front of the observer, the formal description of brain activity, and the visual awareness of the observer. These are three disjunct ontological levels, so the very notion of "causal connections" is problematic. Some silent assumptions in current accounts indeed invoke "magic", we identify internal and external local sign as instances. We attempt an account of the shading cue that avoids at least some of these pitfalls. We conclude that (for the human observer, machine vision has different objectives) the shading cue allows "direct perception" of surface shape.

2.1 Structure of the Scene in Front of an Observer: Radiometry

The radiometric problem of "Shape From Shading" is simple in principle, but frequently intractable in practice [3, 7, 12, 16, 57]. Consider the simplest case.

In the simplest setting, one considers a surface being illuminated with a uniform, unidirectional beam. This constraint may obtain in real life, for instance,

Jan Koenderink
Laboratory of Experimental Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, 3000 Leuven, Belgium
e-mail: jan.koenderink@ppw.kuleuven.be
Universiteit Utrecht, Faculteit Sociale Wetenschappen, Afdeling Psychologische Functieleer, Postbus 80140, 3508 TC Utrecht, The Netherlands
e-mail: J.J.Koenderink@uu.nl

Andrea van Doorn
Universiteit Utrecht, Faculteit Sociale Wetenschappen, Afdeling Psychologische Functieleer, Postbus 80140, 3508 TC Utrecht, The Netherlands
e-mail: A.J.vanDoorn@uu.nl
Department of Industrial Design, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands
e-mail: A.J.vanDoorn@tudelft.nl

G. Citti and A. Sarti (eds.), *Neuromathematics of Vision,*
Lecture Notes in Morphogenesis,
DOI: 10.1007/978-3-642-34444-2_2,

direct sunlight is a good approximation [18, 30, 43]. (However, it is the worst setting for effective shading, overcast sky being much more useful, as any professional photographer knows.) The surface is supposed to be uniform. This may obtain in real life, at some reasonable scale. For instance, a blank piece of paper will do at the millimeter scale. The BRDF (Bidirectional Reflectance Distribution Function [47]) is supposed to be constant. This is the so called "Lambertian assumption" [12, 41]. Although such surfaces don't exist, the constant may obtain approximately, if the range of surface spatial attitudes is not too large. For instance, a piece of blotting paper is a good approximation for angles of incidence and observation not too far from normality. Vignetting is supposed to be absent. "Vignetting" indicates occultation of the source by parts of the object [27]. Thus, various parts of the object are illuminated by mutually different "effective" sources. An example is an apple seen under overcast sky illumination. The constraint can be met in many cases, for instance direct sunlight away from the attached shadow boundary. Multiple scattering is supposed to be absent. That constraint can only be met for convex objects, which is very restrictive [27]. However, a surface that is fairly flattish yields a good approximation [27].

Notice that these constraints are quite limiting in their totality. However, the constraints are automatically met if you sufficiently limit the region of interest (ROI). Such a constraint serves to select approximately homogeneous and flattish surface patches. However, the "effective" source might well be quite different from the nominal one. For instance, it could be modulated by vignetting and/or multiple scattering effects. If one considers smaller ROIs the problem becomes simpler, but the effective direction of illumination becomes more variable.

This makes it likely that biological "shape from shading" [8–10, 45, 56] will be limited to rather smallish ROI's, and makes it likely that methods that do not explicitly require knowledge of the effective direction of illumination will be preferred.

2.1.1 Shading Geometry

In the simplest case, the radiance incident upon the eye is independent of the viewing direction. It depends only on the structure of the incident beam. In the simplest case we can summarize the incident beam by a "light vector", which is a suitable average over the directions represented in the beam [36–38, 46]. The radiance scattered to the eye is proportional to the irradiance caused by the incident beam. If the BRDF is approximately constant, the angle of incidence is crucial. The radiance scattered to the eye is proportional to the illuminance, which is proportional to the cosine of the angle of incidence. This implies that the scattered radiance as a function of location is approximately proportional to the change of spatial surface attitude [28, 29, 31, 33, 53]. Notice that there are numerous complicating factors whose influence we have somehow ignored. The degree to which these approximations "work" depend all on the size of the ROI. They also depend upon the form of the BRDF, vignetting, and so forth, something we will ignore from here on, but should not be forgotten.

2.1.2 Outline of the Paper

In order to proceed, we need to connect brain activity to scene geometry, to physics, and to visual awareness. We will do this in steps. First we discuss the nature of the relevant brain activity. Then we discuss the relevant nature of the "scene", which involves merely the local surface shape landscape. Then we have to connect these (ontologically distinct) universes in some way. Finally, we need to establish the relation to visual awareness. Needless to say, this is a very ambitious program, and we are unlikely to succeed. The goal of the exercise is mainly to obtain a more focussed conceptual grip on the problem.

2.2 Visual Front-End

Visual awareness is perhaps best understood as a "user interface" [25]. The user interface depends mainly on templates of various sorts. Microgenesis tries templates by running "reality checks" against the front end neural activity. As a result, such "hallucinations" may gain any desired degree of objectivity.

The "reality checks" are against the activity of the "visual ("optical" would have been more apt) front end". It is hard to define the boundaries of the front end. Here we simply talk of V1, the primary visual cortex. However, we are prepared to extend this definition when opportune, neuro-anatomy proper not being our primary interest.

The visual front end is a volatile buffer, that is continually being overwritten by the world (the radiance at the corneas). It is in many respects like the beach, which "represents" footprints. Wet (but not too wet) beaches are better than dry beaches in that respect. Likewise, the front end is optimized to retain useful structure, and discard useless junk. Moreover, it has a number of limiting constraints. For instance, at any location, the front end optimizes dynamic range by adjusting the local set point and gain. These are not forwarded to more remote parts of the brain. The local set point and gain are simply lost. Only in very local regions can one count on the gain to be constant. An example might be a "column", but we can't be sure.

2.2.1 Brain Activity due the Optical Structure

One simple principle used by the front end is to prefer derivatives of some useful degree over absolute values [24]. This is useful (at least in principle), because it obviates the need for absolute calibration. The generic example is the Laplacean ("Mexican hat receptive field"), which encodes the difference between a local region and a larger one. Another principle is to prefer ratios over absolute values. The ratio of two values of a non-negative quantity is evidently independent of the absolute calibration. A combination of these principles is to retain the "contrast gradient". The "contrast gradient" may also be defined as a "logarithmic derivative" (same thing). For a retinal illuminant $I(x,y)$ (Cartesian spatial coordinates $\{x,y\}$), the contrast gradient is (defining $I_0 = I(x_0,y_0)$):

$$\mathbf{C}(x_0, y_0) \propto \frac{1}{I_0} \nabla I(x,y) = \nabla \log I(x,y) \tag{2.1}$$

Such a contrast gradient might be available in the visual front end, at least in regions of limited extent. Over larger regions one runs into problems because the location dependent gain factors cannot be assumed to be available.

The contrast gradient is a vector quantity. The front end represents it in terms of an overcomplete basis of first order directional derivatives. The kernel of such a derivative in the x-directions is [22]

$$E(x,y,s) = x \frac{\mathrm{e}^{-\frac{x^2+y^2}{2s^2}}}{2\pi s^4}, \tag{2.2}$$

where the parameter s parameterizes the scale of the derivative operator. In polar coordinates $\{\rho = \sqrt{x^2+y^2}, \varphi = \arctan y/x\}$, the derivative in the direction φ_0 is

$$E(\rho, \varphi, s, \varphi_0) = \rho \cos(\varphi - \varphi_0) \frac{\mathrm{e}^{-\frac{\rho^2}{2s^2}}}{2\pi s^4} \tag{2.3}$$

The overcomplete basis is indexed by φ_0, it is a "cortical column" of "edge detectors". The contrast gradient is represented through the total activity in the column. It is the first Fourier component of the activity. All other Fourier components may be ignored. They represent noise. The overcomplete basis has the advantage of yielding a very robust representation, and not forcing you to decide on a fiducial coordinate frame.

Notice that the edge detector kernel depends upon an additional parameter s. This is a nonnegative quantity, that represents the *scale*. The edge detector yield exact first order directional derivatives *at some finite scale* [22]. We assume that the brain has a range of scales available. Thus, the topic of "scale space" is of basic importance.

Both the size of the ROI, and the scale of computation, will depend upon the task at hand. The square of the ratio of ROI diameter and scale value represents the number of degrees of freedom in the local representation. We expect it to be typically limited, not exceeding a small "icon" (as familiar from computer desktops) in complexity.

2.2.2 *Representation of the Optical Structure*

In the previous subsection we discussed the activity of the front end "at a point". This is obviously not sufficient for our goal. The contrast gradient (modulo gain) at a point is useless. What is needed is the *spatial variation* over a region of (at least approximately) constant gain. We need a contrast gradient *field*.

2.2.2.1 Local Sign

The first problem one meets here is that of *local sign*. How do brain "algorithms" take account of location? This is a fundamental problem, first noticed by Lotze [42], that is conventionally ignored in contemporary accounts. Many people apparently believe that somatotopy renders the problem a non-problem. This is naive. It is enough to consider a thought experiment in which a super-surgeon carefully permutes V1 cells, taking the utmost care to leave all mutual connections intact. Will this produce a local deformation of the visual field? If you don't count on magic, your answer has to be no. The "machine" is still the same, but the somatotopy has been destroyed.

There is one way to deal with local sign locally. One designs a complex receptive field that implements the complete algorithm. This takes care of the problem. Local sign has been "encapsulated". An "edge detector" is the simplest instance [12]. It is really a bilocal entitity, wrapped up into a purely local one.

In general, the local sign problem has never really been solved. The problem is too hard to tackle in this paper. We will simply ignore it, but we will acknowledge the problem, and use it to change our treatment to rather formal, and abstract, whenever the local sign issue arises. There seems to be little use in pretending to suggest "neural implementations" when major basic problems are left open. Of course, our formal treatment will be implementable in principle once ideas concerning local sign take more rigid form.

2.2.2.2 The "Contrast Gradient Image"

The "contrast gradient image" is a formal description of the structure of V1 activity that depends upon the local sign issue. Thus we treat it formally, instead of suggesting some "neural network" implementation. We make the essential role of local sign explicit.

The contrast gradient field is a map of part of the visual field (this is where local sign comes in) to the space of possible gradient values. Both spaces are two-fold extended. The patch of the visual field is a topological disk, say. A convex region will be most convenient. The "gradient space" is a vector space, with a well defined origin. We will mostly consider a disk centered on that origin. The diameter of the region depends upon the dynamic range of the edge detectors.

The map is "from the plane into the plane" [58], thus its generic singularities will be folds and cusps. Non-generic cases will no doubt occur, mostly of a degenerate kind. For instance, a region of uniform illumination will map on the origin, thus the map will collapse on a point. We will mainly be interested in the generic case.

The singular entities are curves (the folds) and points (the cusps) [1]. (See figure 2.1.) The configuration of singular elements alone suffices to characterize the contrast gradient image in a qualitative (or, rather, semi-quantitative) way. We will consider this contrast gradient image as the "representation of the optical structure" in the visual front end. It contains all that is of relevance to the shading cue ("cue" is a notion due to Berkeley [4]), and nothing else.

Folds occur when the Jacobian of the image vanishes, that is to say when $\frac{\partial C_x}{\partial x}\frac{\partial C_y}{\partial y} - \frac{\partial C_x}{\partial y}\frac{\partial C_y}{\partial x} = 0$. This implies $I_{xx}I_{yy} - I_{xy}^2 = 0$. This occurs on curves. Cusps occur when the gradient of the Jacobian is tangent to the curve, they occur as isolated points [1].

The contrast gradient image is a data structure that makes the spatial variation of the local edge detector activities *explicit*. It is necessary, because the raw presence of a distribution of edge detector activities in the brain only contains this structure *implicitly*. Because of the retinotopic structure of the brain it is visible to an external observer of V1. For instance, it might be revealed through some smart imaging technique. However, it is not available to V1 itself, *because V1 cannot "see itself"*. It is a sufficient summary of what might be "seen" (that is to say, made explicit as some data structure) by some "higher" part of the brain.

The very concept of "contrast gradient image" depends upon the existence of a local sign. The more primitive apparatus (the edge detectors) is simple and well understood. The local sign is not. We will not speculate on the form the contrast gradient image might take in terms of higher brain activity.

2.3 Local Shape

The description of "local shape" has nothing to do with the brain *per se*. It is a geometry of certain aspects of the scene in front of the observer. It does not even specify the optical structure impinging on the eyes directly. With local shape we mean the "curvature landscape" of the boundary surfaces of environmental objects. "Local" implies that we study the curvature in the neighborhood of a point. Of course, a certain scale is always implied, since there exist no smooth environmental surfaces if the scale is left unspecified. A local, smooth surface, is an entity that can be described sufficiently well by way of a Taylor expansion up to some reasonable order (say less than ten). This implies both a scale and an extent [23]. One typically has a choice here. For instance, a treetop can be treated as a surface on one scale, but

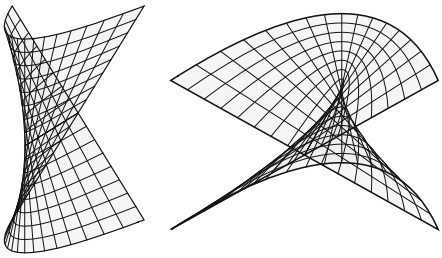

Fig. 2.1 Examples of the generic singularities of the contrast gradient image. At left a fold, at right a cusp. Near a fold gradient space is either covered zero or twice, near a cusp once and thrice.

not on the scale where individual leaves appear. This type of description has been intuitively used in the visual arts for centuries.

A local surface element can be located by its visual direction and distance. It has a spatial attitude that may be specified by its slant and tilt with respect to the visual direction and the vertical. This specifies the surface element as a "planelet" in the sense of Barrow. The deviations from the tangent plane may be denoted "surflets". In the lowest relevant order a surflet is described by its *curvature*, a "shapelet". The curvature varies from point to point, one has a "curvature landscape". The formal description is simply the classical differential geometry of Euclidean space as pioneered by Gauss.

Perhaps unfortunately, the classical theory is not particularly fit to describe the geometry as relevant to a stationary, monocular observer. We develop the necessary geometry in the next subsection.

2.3.1 The Geometry of "Pictorial Space"

When you look *at* a painting you are visually aware of a flat object, embedded in Euclidean space (that is the "space you move in"), covered with pigments in some simultaneous arrangement. When you look *into* the painting (we assume a "realistic" rendering, say a generic late nineteenth century landscape painting), you become aware of a "pictorial space". This pictorial space is fully detached from the space you move in. For instance, both your eye and the picture surface are in the space you move in, but neither of them is to be found in pictorial space. The space you move in and pictorial space don't even meet in the picture surface (as the world and the reflected world do in a mirror surface).

The geometry of pictorial space has been extensively researched, and formal accounts with excellent predictive power exist [32–35]. We use this formalism here, as it applies equally to pictorial space and to the visual space of a stationary, monocular observer.

Pictorial space has the structure of a fiber bundle [17, 21, 33], namely the depth domain over the visual field. For simplicity, we describe the visual field as a Euclidean plane $\mathbb{E}^2$. As a convenience, we fit it with a Cartesian coordinate frame, we denote the coordinates $\mathbf{r} = \{x, y\}$. The origin is arbitrary, for convenience we place it at the center of the ROI. The depth domain (parameterized by the z–coordinate) has the structure of the affine line $\mathbb{A}^1$. There is no origin of the depth domain, since "absolute depth" is a non-entity, and there is no preferred scale. Thus, pictorial space has the structure $\mathbb{E}^2 \times \mathbb{A}^1$. In the simplest cases depths on different fibers are coordinated with a global, linear gauge field. The gauge can be geometrically represented by two parallel, planar cross sections, one defining an arbitrary origin, the other an arbitrary unit point, on each fiber. The gauges are idiosyncratic and often change over time, even for a single observer. The group of similarities ("proper movements" for $\eta = \gamma = 1$)

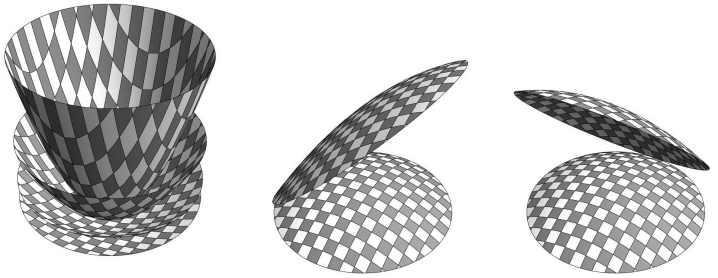

Fig. 2.2 Action of similarities on spheres of the second kind in pictorial space. At left the action of similarities of the second kind (parameter γ in eq. 2.4). At center and right the action of isotropic rotations (parameters ρ_x, ρ_y in eq. 2.4).

$$x' = \eta(x\cos\varphi - y\sin\varphi) + \tau_x, \tag{2.4}$$

$$y' = \eta(x\sin\varphi + y\cos\varphi) + \tau_y, \tag{2.5}$$

$$z' = \rho_x x + \rho_y y + \gamma z + \tau_z, \tag{2.6}$$

describes gauge transformations [33,49,54,59]. It is an 8-parameter group, whereas the corresponding group in Euclidean $\mathbb{E}^3$ is only 7-parameter. We will henceforth set $\eta = 1$, $\varphi = 0$, and $T = \{\tau_x, \tau_y\} = 0$, that is to say, ignore the Euclidean transformations in the visual field. The remaining transformations are a parabolic rotation, parameterized by $R = \{\rho_x, \rho_y\}$, a "similarity of the second kind" parameterized by γ, and a depth translation, parameterized by τ_z. (See figure 2.2.) The translation merely shifts the gauge planes along the fibers, and is generally irrelevant. We ignore it here. The parabolic rotation affects the spatial attitude of the gauge planes, and the similarity affects their spacing.

A "(depth) relief" is a cross section of the fiber bundle. It can be specified as a depth map $\{x, y, z(x,y)\}$. We will consider depth maps *modulo* arbitrary gauge transformations. This describes the nature of pictorial reliefs in considerable quantitative detail.

The geometry of pictorial space is no doubt due to the fact that the optical structure at the eye specifies the scene in front of the observer only partially. The gauge transformations describe the generic ambiguity for many "depth cues". Consider the shading cue for instance. Suppose the ROI is filled with a uniform illuminance. Could it be due to an illuminated surface in the scene? Sure, it could, although this is not necessarily the case. Suppose it is, what may one infer with regard to the shape of the surface? Well, if the illumination is uniform, then (in the generic case), the surface has to be *planar*. Notice that *any plane will do*. Thus the set of possible inferences is simply $z(x,y) = 0$, *modulo* arbitrary gauge transformations.

2.3.2 *Differential Geometry of Pictorial Space*

The differential geometry of pictorial space is similar to, but different from, the familiar differential geometry of Euclidean space $\mathbb{E}^3$ [20, 33, 49, 54, 59]. Consider the metric of $\mathbb{E}^2 \times \mathbb{A}^1$ induced by the gauge transformations. The Euclidean distance in the visual field is conserved, it may be used as the metric of pictorial space. Notice that this renders all points on a single fiber as coincident. The fibers are isotropic (null-)directions. Two points on a single fiber can be assigned a "special distance", which is also conserved. However, the special distance applies only to such "parallel points". The angle measure in the visual field is elliptic, just the familiar (periodic) Euclidean angle. In an isotropic plane, the angle measure is parabolic, thus not periodic. It is measured as the arc length of a "unit circle of the first kind". For instance, in the plane $y = 0$, the unit circle with center at the origin consists of the lines $x = \pm 1$. Thus the slope of the line from $\{0,0,0\}$ to $\{x,0,z\}$ is simply z/x.

A regular plane is a planar cross section. Thus, it does not contain an isotropic direction. The fibers meet this plane orthogonally, that is to say, the isotropic angle is infinite. Thus the isotropic direction is the normal of *any* regular plane. The implication for differential geometry is that the concept of "surface normal" cannot play the dominant role as it does in the conventional treatments of Euclidean differential geometry. One uses the tangent planes instead. The tangent planes can be parameterized by their slopes, that is their depth gradient $\{z_x, z_y\}$. The map of a cross section to its gradients, is the "gradient image" of the relief. It can be regarded as the isometric stereographic projection of the "spherical image" of the relief. The spherical image of a surface is a map of the surface on the "unit sphere of the second kind" $\{x, y, (x^2 + y^2)/2\}$ by parallel tangent planes. The stereographical projection maps $\{x, y, (x^2 + y^2)/2\}$ on $\{x, y, 0\}$ (the "center" of the sphere is $\{0, 0, \infty\}$). The stereographical projection is evidently isometric, not just conformal. Notice that this is analogue to the construction of the Gaussian normal spherical image in Euclidean differential geometry. Gauge transformations simply translate and/or magnify the spherical image. Thus, the relative metrical structure of the spherical image defines the curvature landscape.

Near the origin we may transform any relief to the form

$$z(x,y) = \frac{1}{2!}\left(a_{xx}x^2 + 2a_{xy}xy + a_{yy}y^2\right) + \mathrm{O}[x,y]^3, \tag{2.7}$$

by adopting a suitable gauge. This is the shapelet descrion we use in this chapter. The differential invariants $K = a_{xx}a_{yy} - a_{xy}$ and $2H = a_{xx} + a_{yy}$ then define the Gaussian and the mean curvatures [15, 23]. Notice that these expressions are much simpler than the corresponding expressions for the Euclidean case. Reason is the absence of "foreshortening".

2.3.2.1 Shapelet Space

The deviation from planarity near a point is denoted the "surflet' at that point, adapting Barrow's formalism [5, 11]. (See figure 2.3.) The lowest non-trivial description

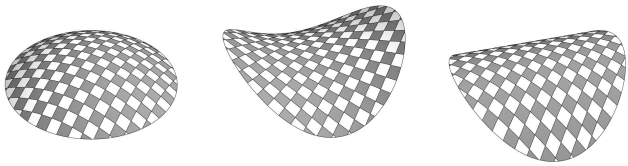

Fig. 2.3 "Surflets" can be scaled and added. Here the umbilical surflet at left and the saddle-shaped surflet at center are added so as to yield the cylindrical surflet at right.

is in terms of the second order terms in a Taylor expansion about the origin. In the geometry of pictorial space that implies equation 2.7. A shapelet may be parameterized by the coefficients $\{a_{xx}, a_{xy}, a_{yy}\}$, and indicated as a point in "shapelet space". This is useful, because it allows us to regard "curvature landscapes" as surfaces (maps of the visual field, thus immersions) in shapelet space.

The parameterization by $\{a_{xx}, a_{xy}, a_{yy}\}$ is not very useful, because referred to the arbitrary Cartesian frame. One may do better by rewriting the form as

$$r\frac{x^2-y^2}{2}+sxy+t\frac{x^2+y^2}{2}, \tag{2.8}$$

where

$$r = \frac{1}{2}\left(a_{xx}-a_{yy}\right), \tag{2.9}$$

$$s = a_{xy}, \tag{2.10}$$

$$t = \frac{1}{2}\left(a_{xx}+a_{yy}\right), \tag{2.11}$$

the point being that the shapelet $(x^2+y^2)/2$ is rotationally invariant, whereas the shapelets $(x^2-y^2)/2$ and xy transform as a pair under rotations of the Cartesian frame. Thus, t and $w=\sqrt{r^2+s^2}$ are differential invariants, whereas $\varphi = \frac{1}{2}\arctan s/r$ describes the "orientation of principal curvature" with respect to the Cartesian frame. It may be called the "attitude" of the shapelet, whereas the ratio $w:t$ describes its shape *proper*, and $\sqrt{t^2+w^2}$ its *amount* of curvature.

The "Casorati curvature" [6] $C=\sqrt{r^2+s^2+t^2}$ can be interpreted as the R.M.S. deviation from planarity (simply defined through a suitable limiting process), or also as its R.M.S. sectional curvature, or (again equivalently) $\sqrt{(\kappa_1^2+\kappa_2^2)/2}$, where $\kappa_{1,2}$ are the "principal curvatures": in a frame rotated to set $\varphi=0$ the shapelet is described as $(\kappa_1 x^2+\kappa_2 y^2)/2$. Vanishing planarity implies $C=0$, thus for a proper shapelet one has $C>0$.

The parameter

$$\sigma = \arctan\frac{\kappa_2+\kappa_1}{\kappa_2-\kappa_1} = \arctan\frac{t}{w}, \tag{2.12}$$

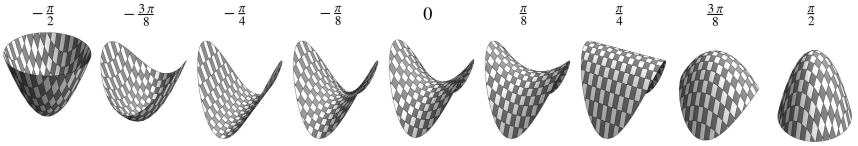

Fig. 2.4 The shape index series of quadric surflets. Notice that the umbilicals are at the endpoints. The symmetrical saddle at zero is congruent to its own mould. Shapes of opposite signs are related as object to mould.

(where we use the convention $\kappa_1 \geq \kappa_2$) is a pure shape descriptor, the "shape index" [23]. It assumes values in the range $[-\pi/2, +\pi/2)$. (See figure 2.4.) At the endpoints of the range the shapelet is "umbilical", that is rotationally symmetric, so the orientation φ is not defined. At $\sigma = 0$ the shapelet is a "symmmetrical saddle", meaning that its inverse (inverted depth) is congruent to itself. For non-zero shape index the shapelet and its inverse are in the relation of a shape to its mold.

One has

$$t = C \sin \sigma, \tag{2.13}$$

$$r = C \cos \sigma \cos 2\varphi, \tag{2.14}$$

$$s = C \cos \sigma \sin 2\varphi, \tag{2.15}$$

thus $\{r,s,t\}$ and $\{C,\sigma,2\varphi\}$ are natural Cartesian and polar coordinates of "shape space". The space is naturally polarized by the r-direction. The line $r = s = 0$ contains umbilicals, and the principal directions are undefined on it. The right circular cones of semi-top-angle $\pi/4$ with this line as axis are the locus of cylindrical shapelets. Inside the cones one finds hyperbolical (saddle-like) shapelets, outside elliptical ones (either like the outside, or like the inside of egg shells).

Because absolute size is largely irrelevant in vision, it is natural to define a Riemann line element [52]

$$\frac{\mathrm{d}r^2 + \mathrm{d}s^2 + \mathrm{d}t^2}{r^2 + s^2 + t^2} = \mathrm{d}\mu^2 + \mathrm{d}\sigma^2 + 4\cos^2\sigma\,\mathrm{d}\varphi^2, \tag{2.16}$$

(where $\mu = \log C$) as a natural metric for shape space. The geodesics are planar logarithmic spirals in planes through the origin (of course only arcs contained in a half-space are relevant). In this metric the shape index scale (for constant Casorati curvature and orientation) is linear, so is the log-Casorati curvature (for constant shape index and orientation) scale, and so is the orientation (for constant shape index and Casorati curvature) scale. On spheres of constant Casorati curvature the spherical distance scaled by log-Casorati curvature is a geodesic distance.

2.3.2.2 Curvature Landscapes

A "curvature landscape" is a field of shapelets. We can represent it as a map of the ROI in the visual field into shape space. It will generically be an immersed

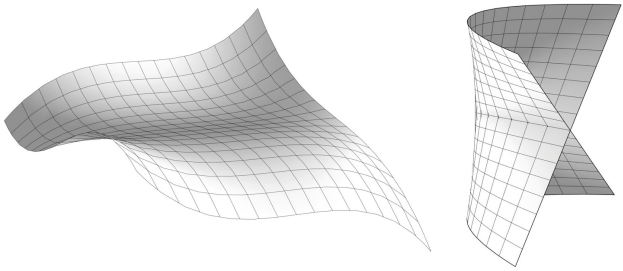

Fig. 2.5 Examples of a "Whitney umbrella". At left the surface, at right the curvature landscape in shape space.

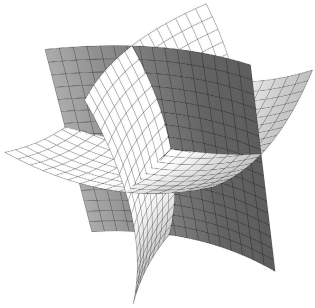

Fig. 2.6 A triple point

surface. Such immersions tend to have only mild singularities (there is lots of room in the space), generically they are Whitney umbrellas [14] (or cross-caps) (figure 2.5) and triple points (figure 2.6). Whitney umbrellas occur when two relations between the cubic terms in the Taylor expansion are simultaneously satisfied, thus at isolated points (the condition is $(a_x)_{xx}(a_x)_{yy} - (a_x)^2_{xy} = 0 \wedge (a_y)_{xx}(a_y)_{yy} - (a_y)^2_{xy} = 0$). Because shading is proportional with the slope of the surface in some direction, it implies $I_{xx}I_{yy} - I^2_{xy} = 0$, thus a point on a fold of the contrast gradient image, when the surface is illuminated.

Notice that not just any immersed surface represents a curvature landscape, in order to classify, the so called "integrability equations" (in the Euclidean case these are the Codazzi-Mainardi equations) have to be satisfied. We are confronted with a Pfaffian problem. For instance, it is evidently required that $\partial a_{xx}/\partial y = \partial a_{xy}/\partial x$, and $\partial a_{yy}/\partial x = \partial a_{xy}/\partial y$. This means that there is a constraint on curvature landscapes, and we cannot simply apply the generic taxonomy of singularities.

In one experiment we generated hundreds of high order, random polynomial surfaces. One finds Whitney umbrellas galore. Triple points are much harder to find, but that is because one has to search over large regions, they are not local phenomena. They certainly occur, because one may always start with three intersecting planes and assemble them into a single surface.

The notion of a "curvature landscape" is necessary if the local description stops at the second order. Of course, similar notions will still be necessary if one includes the cubic and quartic structure in the local description. The curvature landscape yields a global data structure in terms of a map. This is similar to a geographic description that mentions the relative locations of local features like mountains, lakes, and so forth. ("Mountain range" being a simple example.) Without such a global map one has only a bag of mutually unrelated features. In our case the "glue" is what might be called "external local sign", an awareness of the directions in the space external to the eye corresponding to retinal locations.

2.4 The "Shape From Shading" Problem

The "Shape From Shading Problem" can be framed in a large variety of ways [12]. Most of the conventional settings are hardly relevant to human (or animal) vision [2, 4, 26, 30, 40, 43, 48, 50, 51, 53, 55]. Here we impose the following a priori constraints on the matter:

— the "data" are the contrast gradient image, based upon the edge detector activity of the visual front end;
— the desired inference is a curvature landscape, that is the inverse of the image of the ROI in shape space.

Shape space is not a brain activity, or anything like that, it is a formal construction used in microgenesis.

The "microgenesis" of visual awareness is a hypothetical pre-conscious process that generates awareness. From experimental phenomenology we know that microgenesis is a systolic process that regenerates awareness continuously, a single "beat" taking less than a tenth of a second. The process generates hypotheses (or "hallucinates") and runs reality checks against visual front end activity. In a single beat it launches a volley of threads (representing different hypotheses), that may be novel, but typically are diversified threads from the previous beat. Some threads from the previous beats might be terminated. A competition between the threads leads to a final "winner" that enters visual awareness. Thus, momentary awareness is the result of an evolutionary process that runs on a very short time scale. In the generation of each volley current situational awareness ("gist") and goals (input from cognition and emotional states) play a role, though the process is largely autonomous.

"Shape from shading" starts when the microgenetic process selects "shaded surface" as a hypothesis. The same retinal illuminance may be interpreted in many different ways. "Shaded surface" is just one. (For instance, "painted flat picture" is another, "non-uniformly illuminated surface" yet another, and so forth.) It will typically involve a number of mutually related hypotheses, some aimed at material properties, others at the light field, etc. Here we concentrate on the "shaded surface" hypothesis. The act of generating a "shaded surface" hallucination serves to turn the front end structure into (meaningful) "data". The meaning derives from the hypothesis.

2.4.1 Naive Radiometry

The microgenetic process may assume very little about the geometrical layout, and physics of the scene. All it can do is assume an "uniformly illuminated surface", which involves a number of associated hypotheses, all of which might become falsified in reality checks at a number of levels, from front end activity to reflective thought. These include:

— the surface is a smooth Lambertian surface (no space-variant pigmentation, no specularities inside the ROI);
— the illumination is oblique (no frontal illumination);
— the surface is illuminated throughout (no attached shadow terminator in the ROI).

Full analysis of these assumptions and their interrelations would take many pages, we skip it here.

Notice that these are no detailed assumptions concerning surface attitude, nor about the light field. The shading is subject to the well known "bas-relief ambiguity". What this all implies is that the magnitude of the contrast gradient cannot be distinguished from the obliquety of the illumination. The structure of the contrast gradient field depends upon the local shapelet and the tangential component of the light vector over the surface ("surface illuminance flow"). Let the shapelet be parameterized as in equation 2.7, and let the direction of surface illuminance flow as projected in the visual field be $\{\cos\vartheta, \sin\vartheta\}$ in the xy-plane.

In order to simplify the formalism we set (arbitrarly, but without loss of generality) $\vartheta = 0$. Then the contrast gradient will (to a good approximation) be proportional with

$$C(x,y) \propto \{a_{xx}, a_{xy}\}. \tag{2.17}$$

The constant of proportionality depends on many things (spatial albedo or BRDF variations, obliquity of the light vector, vignetting, ...). If the surface is significantly corrugated, the contrast gradient will be mainly determined by the curvature landscape though.

2.4.2 The Shape Cue Inference

Notice that the "data", that is the contrast gradient image based upon the front end edge detector activity, is a projection of the curvature landscape in $\{a_{xx}, a_{xy}, a_{yy}\}$-space obtained by dropping a_{yy}, that is the second derivative of the depth in the direction orthogonal to the illumination flow direction in the visual field. Since shape space is a simple linear transformation of $\{a_{xx}, a_{xy}, a_{yy}\}$-space, we find that microgenesis finds a projection of the shape image in the front-end data, and is not in need of any further computation. (See figures 2.7 and 2.8.) There is even a check on the viability of the hypothesis: simply find the curl of the gradient field, if it vanishes the hypothesis can be upheld.

Notice that $a_{yy} = 0$ implies $r = t$. Thus, the "view direction" of the view into shape space subtends a $\pi/4$ angle with the axis of umbilicals.

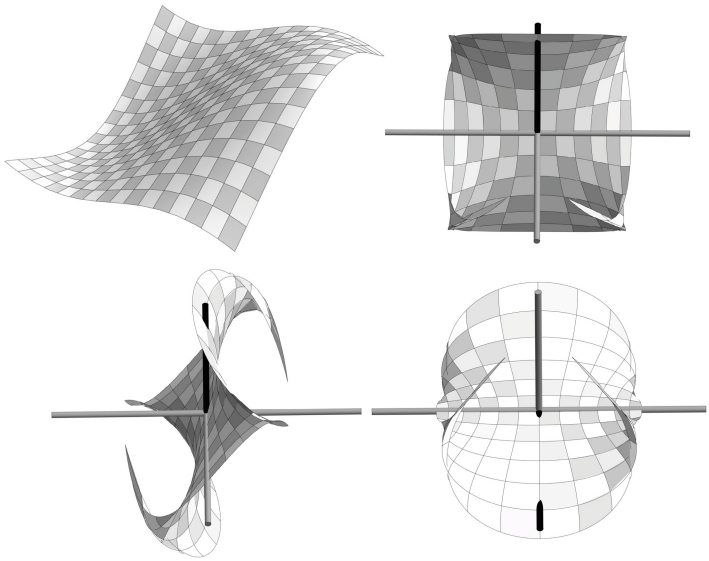

Fig. 2.7 An example of “Shape From Shading”. At top left the habitus of a surface. The other subfigures show three views of the shape landscape in shape (that is *rst*-) space. This surface has no singularities.

One might regard this as an example of “direct perception” as propagated by Gibson [13], in his ecological approach to visual perception. What is conceptually interesting is that the shape from shading problem becomes formally identical to the “shape from movement” problem. One observes a two-dimensional projection of a surface immersed in a three-dimensional space, and attempts to make inferences about the immersed surface (e.g., its projections as viewed from other directions). Only the space is different, it is the “space you move in” in the case of shape from movement, and the space of shapes in the case of shape from shading. It is hard to see that this should make much difference to the brain: these spaces are just as “abstract” as seen from the brain’s perspective. Both play some role in certain perception-action cycles.

The shape image has generically fold and cusp singularities. The projection generically generates folds and cusps in the contrast gradient image. The latter might be called “spurious”, since they depend upon the direction of the projection, whereas singularities in the gradient image that derive from the singularities in the shape image have intrinsic meaning.

Variation of the (relative) illumination direction will clear up such ambiguities. Such variations could be due to movements of the light source (relatively rare), or movements of the surface with respect to the source (common). The latter type of variation can often be brought about by manipulation, thus opening a way to active exploration.

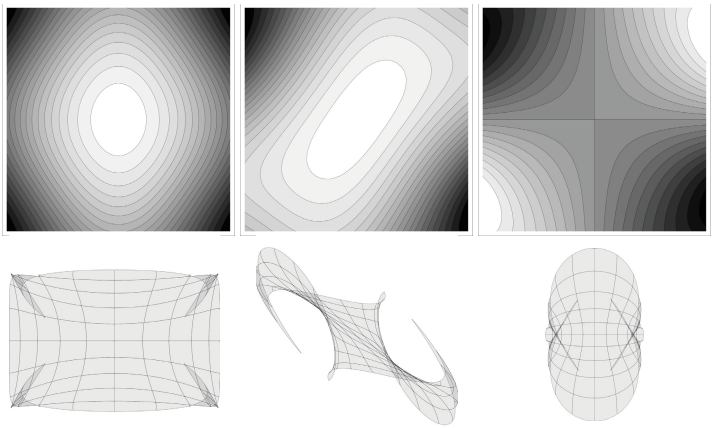

Fig. 2.8 The example from figure 2.7. At top three shadings of the surface for different illumination directions. At bottom the corresponding contrast gradient images. Compare these to the projections of the curvature landscape: they are the same, except for a $\cos\frac{\pi}{4} = \frac{1}{2}\sqrt{2}$ foreshortening.

A final point of some interest is the relation to the "Shape From Shading Problem" as it is usually framed in computer vision with the present setting. One difference is that computer vision frames the problem in Euclidean terms, which introduces some algebraic complications. More interestingly, in computer vision one would not be satisfied with a curvature landscape, but would require an explicit representation of a surface. In the case of the human observer the curvature landscape should most likely be regarded as the end result (see below).

The difference is the need of an integration. One has a Pfaffian system of local tangent quadrics and seeks for an integral surface. There will be a solution if the surface integrability conditions are satisfied (the Codazzi-Mainardi equations), which is never exactly the case if the field of quadrics is due to observations. The computer vision methods differ primarily in the way they handle this problem, which has nothing to do with the shading cue *per se*, it is just a problem of numerical analysis. If one has clean data the integration poses no problem (for an $N \times N$-pixel image one has about $2N^2$ equations for N^2 unknowns, it is mere matter to deal with the ambiguity), things start to become interesting in case the data are "dirty", which they always are in real life. Again, this is not an issue of much biological interest.

2.5 The Shading Cue and Visual Awareness: Phenomenology

The shading cue has been studied extensively in experimental phenomenology. The topic is closely related to that of pictorial space in general. How are pictorial reliefs "represented"? We have been able to show empirically that the representation

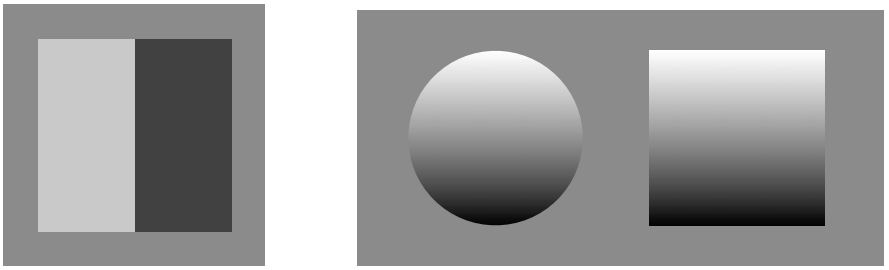

Fig. 2.9 *Left*: The Mach book may be the simplest "shading" stimulus ever invented. Many observers see an open book, or two planar facets meeting in a common edge, subtending a roughly right dihedral angle. Ernst Mach [44] interpreted this as a direct causal connection between intensity and the awareness of spatial attitude. *Right*: The circular disk filled with a linear gradient is the "canonical stimulus" that has been used for almost two centuries in psychophysical shape from shading research [19, 50, 51]. Many observers become aware of a spherical pictorial surface when viewing this stimulus. The contrast gradient image is degenerate (a point). The square contains the same gradient. Many observers see it as a cylinder. *Any* quadric surface could yield this gradient, for instance, a symmetrical saddle is a perfectly valid inference. It is *never* reported.

is not a depth map, but more likely a map of spatial surface attitudes (a depth gradient field), or possibly (this issue is still open) a curvature landscape. The present treatment of shape from shading fits perfectly in this general framework.

Most of the psychophysical work has been concentrated upon very singular cases. The simplest instance is no doubt the "Mach book" (figure 2.9 left), but the most widely used stimulus in shape from shading research is a linear illuminance gradient limited to a circular disk. (See figure 2.9 right.) Thus the contrast gradient image is a point that does not coincide with the origin. Thus, the corresponding curvature landscape would generically be a point. (A line in the direction of projection in the *rst*-space being evidently non-generic.) Hence, the inference would be "any quadric". Observers report convex or concave elliptical shapes, which are indeed "solutions", albeit very specific ones. Hyperbolical shapes are never reported, although they are equally valid inferences. Apparently microgenesis applies additional constraints. In this case it is the shape of the outline (circular). If the outline is square (two edges parallel to the gradient), observers report convex or concave cylindrical shapes (see figure 2.9 right).

One would expect that the next round of empirical research would concentrate on curvature landscapes that are degenerated to curves instead of points. However, we know of no instances. So the next round should involve generic curvature landscapes. Unfortunately, we haven't seen much activity on this topic either (except from some work of our own). The problem of "shape from shading" is pretty much open in the context of experimental phenomenology.

2.6 Conclusion

We have presented a discussion of the shading cue that is decidedly different from its conventional formulation in machine vision. The reason is that the present treatment has been focussed upon the phenomenology of human visual awareness, and upon an understanding of the brain "from the inside" as it were. We have refrained from "representations" that are only available to an external observer. The result is a description that renders the "shape from shading problem" trivial. The inference is essentially identical to the observation. Thus, we end up with a theory of "direct perception". This is not to say that the inference is complete, as it cannot be. The resulting ambiguity is very simple in that one obtains a single perspective of the curvature landscape in shape space. This is much like the visual projection itself: you see only the fronts of objects in the scene in front of you.

The treatment requires the existence of both *local sign* (as defined by Lotze), and *external local sign* [39]. None of these is well understood. There exist a number of theories on the genesis of local sign, and mainly speculations on the origin of external local sign. Recent empirical work has revealed that the structure of external local sign in human observers is surprising, as already intuited by Helmholtz and Kepler.

The connection of our treatment with the experimental phenomenology of human visual awareness is still weak. The reason is mainly the lack of a solid body of quantitative empirical results.

References

1. Banchoff, T., Gaffney, T., McCrory, C.C.: Cusps of Gauss Mappings. Pitman Advanced Pub. Program, Boston (1982)
2. Baxandall, M.: Shadows and enlightenment, 2nd edn. Yale University Press, London (2005)
3. Belhumeur, P.N., Kriegman, D.J., Yuille, A.L.: The bas-relief ambiguity. International Journal of Computer Vision 35, 33–44 (1999)
4. Berkeley, G.: An Essay Towards a New Theory of Vision. Printed by Aaron Rhames, at the Back of DickÃs Coffee-House, for Jeremy Pepyat, Bookseller in Skinner-Rows, Dublin (MDCCIX)
5. Boyd, R.W.: The History of the Calculus and its Conceptual Development. Dover, New York (1949)
6. Casorati, F.: Nuova denizione della curvature della supercie e suo confronto con quella di Gauss. Rend. Inst. Matem. Accad. Lomb. 2(22), 335–346 (1889)
7. van Diggelen, J.: Photometric properties of lunar crater floors. Recherches Astronomiques de lÃObservatoire dÃUtrecht 14, 1–114 (1959)
8. van Doorn, A.J., Koenderink, J.J., Wagemans, J.: Light fields and shape from shading. Journal of Vision 11(3), 1–21 (2011)
9. van Doorn, A.J., Koenderink, J.J., Todd, J.T., Wagemans, J.: Awareness of the light field: the case of deformation. i-Perception 3(7), 467–480 (2012)
10. Erens, R.G.F., Kappers, A.M.L., Koenderink, J.J.: Perception of local shape from shading. Journal of Vision 11, 1–21 (2011)

11. Eves, H.: An Introduction to the History of Mathematics. Brooks/Cole – Thomson Learning, London (1990)
12. Forsyth, D.A., Ponce, J.: Computer vision: A modern approach. Prentice–Hall, Upper Saddle River (2002)
13. Gibson, J.J.: The Perception of the Visual World. Houghton Mifflin, Boston (1950)
14. Gray, A.: The Whitney Umbrella. In: Modern Differential Geometry of Curves and Surfaces with Mathematica, 2nd edn., pp. 311, 401–402. CRC Press, Boca Raton (1997)
15. Hilbert, D., Cohn–Vossen, S.: Geometry and the Imagination, 2nd edn. Chelsea, New York (1952)
16. Horn, B.K.P., Brooks, M.J. (eds.): Shape from shading. MIT Press, Cambridge (1989)
17. Husemoller, D.: Fiber Bundles. Graduate Texts in Mathematics, vol. 20. Springer, London (1993)
18. Jacobs, T.S.: Light for the artist. Watson–Guptill, New York (1988)
19. Kleffner, D.A., Ramachandran, V.S.: On the perception of shape from shading. Perception & Psychophysics 52, 18–36 (1992)
20. Klein, F.: Vorlesungen über nicht–Euklidische Geometrie. Springer, Berlin (1928)
21. Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry. Wiley–Interscience, New York (1963)
22. Koenderink, J.J.: The structure of images. Biological Cybernetics 50, 363–370 (1984)
23. Koenderink, J.J.: Solid Shape. MIT Press, Cambridge (1990)
24. Koenderink, J.J.: The brain a geometry engine. Psychological Research 52, 122–127 (1990)
25. Koenderink, J.J.: Vision & Information. In: Albertazzi, L., Van Tonder, G.J., Vishwanath, D. (eds.) Perception Beyond Inference: The Information Content of Visual Processes, pp. 27–56. MIT Press, Cambridge (2010)
26. Koenderink, J.J., van Doorn, A.J.: Photometric Invariants related to Solid Shape. Optica Acta 27, 981–996 (1980)
27. Koenderink, J.J., van Doorn, A.J.: Geometrical modes as a general method to treat diffuse interreflections in radiometry. Journal of the Optical Society of America 73, 843–850 (1983)
28. Koenderink, J.J., van Doorn, A.J.: Illuminance texture due to surface mesostructure. Journal of the Optical Society of America A 13, 452–463 (1996)
29. Koenderink, J.J., Pont, S.C.: Irradiation direction from texture. Journal of the Optical Society of America A 20, 1875–1882 (2003)
30. Koenderink, J.J., van Doorn, A.J.: Shape and shading. In: Chalupa, L.M., Werner, J.S. (eds.) The Visual Neurosciences, pp. 1090–1105. MIT Press, Cambridge (2003)
31. Koenderink, J.J., van Doorn, A.J.: Local structure of Gaussian texture. IEEE Transactions on Information and Systems 86, 1165–1171 (2003)
32. Koenderink, J.J., van Doorn, A.J.: The Structure of Visual Spaces. Journal of Mathematical Imaging and Vision 31, 117–187 (2008)
33. Koenderink, J.J., van Doorn, A.J.: Gauge Fields in Pictorial Space. SIAM Journal on Imaging Sciences 5(4), 1213–1233 (2012)
34. Koenderink, J.J., van Doorn, A.J., Kappers, A.M.L.: Surface perception in pictures. Perception & Psychophysic 52, 487–496 (1992)
35. Koenderink, J.J., van Doorn, A.J., Kappers, A.M.L.: Ambiguity and the 'mental eye' in pictorial relief. Perception 30, 431–448 (2001)
36. Koenderink, J.J., van Doorn, A.J., Kappers, A.M.L., Pas, S.F., te Pont, S.C.: Illumination direction from texture shading. Journal of the Optical Society of America A 20, 987–995 (2003)

37. Koenderink, J.J., van Doorn, A.J., Pont, S.C.: Light direction from shad(ow)ed random Gaussian surfaces. Perception 33, 1405–1420 (2004)
38. Koenderink, J.J., Pont, S.C., van Doorn, A.J., Kappers, A.M.L., Todd, J.T.: The visual light field. Perception 36, 1595–1610 (2007)
39. Koenderink, J.J., van Doorn, A.J., Todd, J.T.: Wide distribution of external local sign in the normal population. Psychological Research 73, 14–22 (2009)
40. Koenderink, J.J., van Doorn, A.J., Pont, S.C.: The "shading twist," a dynamical shape cue. International Journal of Computer Vision (accepted April 6, 2013)
41. Lambert, J.H.: Photometria, sive, De mensura et gradibus luminis, colorum et umbræ. V.E. Klett, Augsburg (1760)
42. Lotze, H.: Microcosmus: An Essay Concerning Man and His Relation to the World (1856-58, 1858-64), trans. Hamilton, E., Jones, E.E.C.: 4th edn. T. & T. Clark, Edinburgh (1885, 1899)
43. Luckiesh, M.: Light and shade and their applications. Van Nostrand, New York (1916)
44. Mach, E.: Die Analyse der Empfindungen und das verhältniss des physischen zum Psychischen. G. Fischer, Jena (1903)
45. Metzger, W.: Gesetze des Sehens. Verlag Waldemar Kramer, Frankfurt aM (1975)
46. Moon, P., Spencer, D.E.: The photic field. MIT Press, Cambridge (1981)
47. Nicodemus, F.E., Richmond, J.C., Hsia, J.J., Ginsberg, I.W., Limperis, T.: Geometric considerations and nomenclature for reflectance. Technical Report MN-160, U.S. Department of Commerce, National Bureau of Standards (October 1977)
48. Palmer, S.E.: Vision Science: Photons to Phenomenology. MIT Press, Cambridge (1999)
49. Pottmann, H., Optitz, K.: Curvature analysis and visualization for functions defined on Euclidean spaces or surfaces. Computer Aided Geometric Design 11, 655–674 (1994)
50. Ramachandran, V.S.: Perceiving shape from shading. Scientific American 258, 76–83 (1988)
51. Ramachandran, V.S.: Perception of shape from shading. Nature 331, 163–166 (1988)
52. Riemann, G.F.B.: Über die Hypothesen, welche der Geometrie zu Grunde liegen. In: Habilitationsschrift, Lecture Published in Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen, vol. 13 (1868)
53. Rittenhouse, D.: Explanation of an optical deception. Transactions of the American Philosophical Society 2, 37–42 (1786)
54. Sachs, H.: Isotrope Geometrie des Raumes. Friedrich Vieweg & Sohn, Braunschweig (1990)
55. Turhan, M.: Über räumliche Wirkungen von Helligkeitsgefällen. Psychologische Forschung 21, 1–49 (1935)
56. Wagemans, J., van Doorn, A.J., Koenderink, J.J.: The shading cue in context. i-Perception 1, 159–177 (2010)
57. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Optical Engineerings 19, 139–144 (1980)
58. Whitney, H.: On singularities of mappings of Euclidean spaces. I. Mappings of the plane into the plane. Annals of Mathematics 62, 374–410 (1955)
59. Yaglom, I.M.: A simple non–Euclidean geometry and its physical basis. Springer, New York (1997)