The Royal Society

Pictorial relief

By Jan J. Koenderink

Helmholtz Instituut, Universiteit Utrecht, Princetonplein 5, 3584CC Utrecht, The Netherlands

'Pictorial relief' is a surface in three-dimensional 'pictorial space'. It is perceived in single flat pictures and clearly has nothing to do with binocular stereopsis but with the interpretation of image structure in terms of relations in the external world. Ways to perform geometrical measurements in pictorial space are presented and a number of empirical results are reviewed. Applications to the theory of optical instruments aiding human vision are discussed.

Keywords: pictorial relief; vision; viewing devices; perception; depth perception

1. The enigma of pictorial space

The invention of the stereoscope by Brewster (1844*a*, *b*) and Wheatstone (1838, 1852) in the early 19th century effectively put the lid on a scientific can of worms: the fact that one knows from experience that one perceives three-dimensional relations even *monocularly*, or, even worse, that one perceives such relations in *flat* pictorial renderings of real or imaginary scenes. That monocular space perception is an impossibility was forcefully argued by Berkeley (1709). After the invention of binocular stereoscopy his arguments seemed to carry even more weight since this presented a scientifically acceptable alternative. The scientific community managed to ignore the plain facts for at least a century. For instance, the fact that excellent depth is perceived when one combines two identical photographs in a stereoscope was discovered by accident and almost immediately promoted to the waste-bin. The embarrassment was that the percept of a scene in depth in such cases is often *stronger* than for a true stereo pair: in the latter case one often perceives a *coullisses space*, that is to say, the scene looks like a series of flat cardboard cut-outs staggered into depth, whereas in the former case the objects in the scene look truly rounded (Emerson 1863; Claparède 1904; Eaton 1919; Streif 1923; Ames 1925; Schlosberg 1941; Koenderink *et al.* 1994). This effect was denoted 'paradoxical stereoscopy' (paradoxical because it was understood to be impossible) and could only be published in obscure journals. Even today many scientists (for the best of reasons) refuse to believe it until they see it. Today the resistance has faded a little, partly because of the recent achievements of machine vision. It has even become somewhat fashionable to study monocular depth vision and paradoxical stereoscopy in experimental psychology. Yet our empirical knowledge, not to speak of our conceptual understanding of these important processes leave much to be desired. I address both aspects in this paper.

Many animals (probably the majority) lack binocular stereopsis (Pettigrew 1986) because they have evolved *panoramic* vision and lack binocular overlap. It is hard to imagine that such species lack spatial competences; in fact, experience immediately falsifies this notion. Binocular overlap of visual fields is a specialization of predators, the kind that sit and watch their prey and are in need of a dependable range estimate

in order to calculate their jump. A large part of the data exploited by species lacking binocular overlap relies on movement, or dynamic, differential perspective (Gibson 1950). I will ignore this important (also for *homo sapiens*) aspect here though because it doesn't apply to pictorial relief.

When one *looks at* a photograph of a three-dimensional scene one perceives the photograph (as an object) as a flat thing, although one is often barely conscious of the fact. This clearly promotes effective action if one desires to acquire or handle such objects. One also looks *into* the photograph (as a carrier of optical information) and thus looks into pictorial space. This space has a three-dimensional character and is what usually dominates conscious perception. Our view can be checked by the surfaces of opaque (pictorial!) objects much like what happens in regular space (I write 'regular space' for the space we move in). Such surfaces are what one calls *pictorial reliefs*.

Here one encounters a variety of conceptual problems. For instance: What is the intrinsic structure of pictorial space? Do pictorial reliefs have the type of structure one expects of smooth surfaces in regular space (the issue of consistency)? Are pictorial reliefs like the surfaces they depict (the veridicality issue)? Are the pictorial reliefs of various observers similar (the issue of subject dependency)? Are the pictorial reliefs evoked by a single picture for the same observer constant over time (the issue of reproducibility) and do they depend on viewing conditions such as monocular or binocular, normal or oblique viewing (the issue of dependence on viewing conditions)? Answers to such questions have first of all to be provided by experiment, since one is at a loss when asked to predict them from first principles.

(*a*) *Theoretical ideas*

Even in the 19th century many scientists knew from experience that monocular perception of space occurs (von Helmholtz 1896), although the topic was not a popular one. Even today, the dominance of monocular vision in daily life is typically much underestimated. One looked for ways to explain it away. Most theories were essentially based on Berkeley's concept of *depth cues*. Depth cues are arbitrary associations. Berkeley gives the analogy of perceiving anger or shame in the looks of a man: the cue is a shift of spectral radiant power towards the long wavelength region (in modern terms). The arbitrariness was stressed especially because it implied uncertainty, a noncausal inference. A small and a large image of a pair of men might either signify a dwarf and a giant at the same distance or two normal men at different distances.

Nowadays one feels that the arbitrary nature of depth cues is the correct biological view (Riedl 1990): acquisition of any competence (either over evolutionary time spans or during the life of an individual) works like that. However, one sees simultaneously that there is little arbitrariness about the depth cues in another sense (Marr 1982): they represent generic regularities of nature, the causal nexus of the optical interaction involving the physical environment. Thanks to the maturation of the field of machine vision one now has respectable quantitative theories of many of the classical depth cues. Such theories allow machines to take effective action without prior learning of arbitrary associations.

The various depth cues differ with respect to the prior knowledge they require. In practice one relies on the concordance between a variety of depth cues. The richer the

information the less prior information is required to reach a certain level of confidence. There can be little doubt that this is the reason that human observers do well in real life but appear very uncertain under artificial laboratory conditions. Indeed, using the scientifically respectable paradigm of stimulus reduction one easily shows that humans are *quite unable to perceive depth relations*, very much in the Berkelean spirit. (For instance, in the case of the shading cue, see Erens (1991, 1993*a*, *b*).) Although the literature is especially rich in such reports, it remains the case that even monocular humans are hardly handicapped in the normal environment. An embarrassment indeed. We believe that the best scientific intentions have led to sterile and indeed largely irrelevant knowledge here.

2. How to study pictorial relief

In order to study pictorial relief one must start by ensuring that it *is seen at all*. This may appear trite, but I personally fail to perceive pictorial relief in many scientific paradigms that purport to address this very issue! One way to ensure its existence is to take a clear photograph of a real object as a stimulus (a good realistic painting would serve as well). Such a picture, of course, contains multitudinous depth cues, so it is not suitable to study any one cue in isolation as scientists are wont to do. However, this is easily remedied when one studies a series of such photographs in which a single cue (shading say) is parametrically varied: one simply records how pictorial relief covaries with the parametric variation (Koenderink *et al*. 1996*b*). One then studies the cue at an *operating point* set by all other cues. But this is necessary anyway since no cue works 'at the origin'. Somehow this simple technique (frequently used in engineering (Truxal 1955)) still has to catch on in the field of the experimental psychology of perception.

The next thing is to *measure* pictorial relief. One might in principle try to describe it in terms of a depth map, a depth-gradient field, a field of curvature, etc. However, the usual geometric tools (straight edges, compasses, etc.) cannot be applied since the pictorial object does not exist in regular space but only in the consciousness of some observer. This fact has seemed an insurmountable obstacle to a true geometrical characterization. One has typically compromised for categorical judgments (e.g. does the surface look convex or concave?). However, this is not at all necessary. A geometrical measurement typically involves the application of a fiducial object (straight edge, taut wire) and judging the fit. This can be done equally well in pictorial space, for I may apply the picture of a fiducial object to the pictorial surface and judge the pictorial fit. After all we feel confident in the judgment that a man is sitting on a chair from perusal of a photograph. This is evidently a judgment involving the fit of the man's body to that of the chair. One might design a variety of tools and thus probe the pictorial surface in various ways (Koenderink *et al*. 1992).

When one says that a certain method will yield samples of the pictorial surface, one really misrepresents the true state of affairs: the pictorial surface in the consciousness of some person cannot be a scientific fact (since it can never be made public), thus one should not say that the samples are *from the surface*. The samples are the objective facts and the pictorial surface is best defined in terms of the measurement. Then what is in the head need not bother one. That different methods will yield different pictorial reliefs is only natural and one should not ask which one is right, but should inquire into the relations between them. In principle this is not different

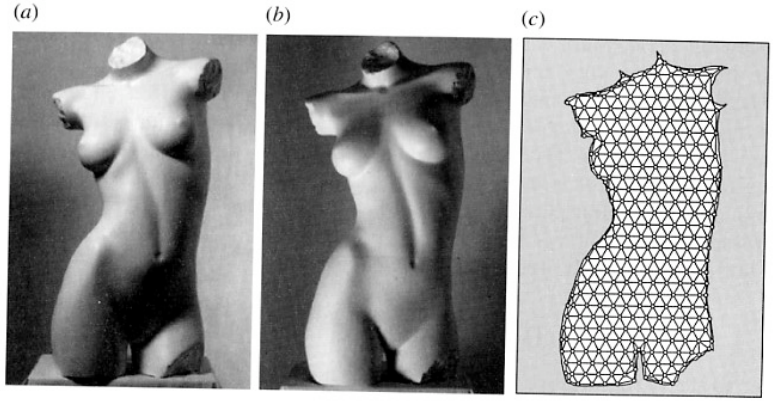

Figure 1. (*a*), (*b*) Two stimuli and (*c*) a triangulation. The stimuli are photographs of the same three-dimensional object, taken from the same camera position. The stimuli are thus geometrically identical. They differ in the shading because the major light source was placed at different positions, top left in the first and bottom left in the second stimulus. Due to the vignetting and multiple scattering this causes complicated global changes in the picture. (*c*) The triangulation that was used in an experiment. Since the same triangulation was used for both stimuli, the results can be compared on a vertex by vertex basis.

from geometry in regular space, although one tends to forget this because experience has taught us that we may count on considerable concordance of results obtained through various methods. One comes to appreciate this point when methods are transplanted to novel and unknown domains (the very small or very large).

(*a*) *The measurement of pictorial surface attitude*

Consider the problem of probing pictorial surface attitude. Since pictorial space naturally factors into a two-dimensional visual field and a one-dimensional depth domain, it is natural to describe surface attitude in terms of (local) slant (Stevens 1983) and tilt. The slant is the angle between the direction of view and the local surface normal (an angle in the range 0–90°), whereas the tilt has to be reckoned with respect to some reference direction in the visual field. For example, one may specify the angle between the vertical and the direction of fastest increase of depth (an angle in the periodic range 0–360°). Since attitude is geometrically represented by the tangent plane, it is natural to apply a fiducial (pictorial) object such as to fit the tangent plane. A suitable object is a *circle*. The picture of the circle will be an ellipse. It will be seen to fit when the ellipse appears like a circle in the (pictorial) tangent plane. Indeed, observers find it natural and easy to make such judgments of fit. We denote this gauge figure 'Tissot's Indicatrix' since Tissot (1887) introduced the ellipse as a graphical indicator of the first order derivative of a surface (in the context of cartography). In the analysis we treat the ellipse as the projection of a circle from which we immediately infer surface attitude (slant and tilt).

The practical implementation is simple (Koenderink *et al.* 1992; Koenderink & van Doorn 1995). One presents the picture (a monochrome photograph) on a CRT screen. The gauge figure is superimposed as a red wireframe image. The observer may adjust

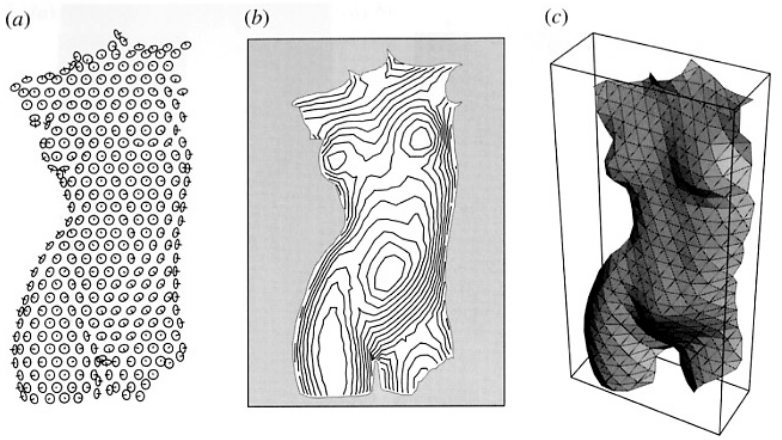

Figure 2. (*a*) The average settings of a subject. Such results can be obtained in about an hour. This figure also shows the habitus of the Tissot gauge figure. In the actual experiment the observer sees only one gauge figure at a time and is not told about the underlying triangulation. The gauge figure appears at the vertices in random order. (*b*) A depth relief map obtained through integration of the data. The stimulus was the first picture of the previous figure. (*c*) A three-dimensional rendering of the relief. Notice that the 'depth dimension' (which is only implicitly present in the stimulus) is made explicit in this rendering.

the shape and orientation of the ellipse (length of the major axis remains fixed) under mouse or trackball control. Bringing about the fit typically takes only a few seconds. Since the ellipse introduces an ambiguity (a minor nuisance) one may add a normal vector to the gauge figure, a line segment erected at the centre, orthogonal to the plane of the circle, with a length equal to the radius. This effectively removes any ambiguity.

For an experiment one first constructs a triangulation of part of the picture plane, covering the surface patch to be measured (figure 1). In practice I prefer a regular hexagonal grid. Practical demands limit the number of vertices to a few hundred, since the subject cannot do more than about 500–1000 settings an hour. The subject never gets to see this data structure, but only sees the gauge figure at one vertex at a time in random order. After all samples have been collected one has several independent attitude settings at each vertex. Thus one may find both the average setting and the spread (covariance ellipse per vertex). The sheer volume of such data exceeds that obtained by classical psychophysical methods by several orders of magnitude. Moreover, the data can immediately be interpreted in geometrical terms, which is hardly ever the case in the classical methods.

I find that the covariance ellipses are very elongated (Koenderink *et al.* 1992; Koenderink & van Doorn 1995), thus the spread is quite anisotropic. Whereas the direction of the local depth gradient is quite well determined, the magnitude is rather imprecise: About 10–20% of its average value (with a certain minimum value of course). It is as if the depth domain is less precisely structured than the visual field.

From the surface attitudes one immediately estimates the depth differences over the edges of the triangulation. Thus one may check surface consistency. The sum

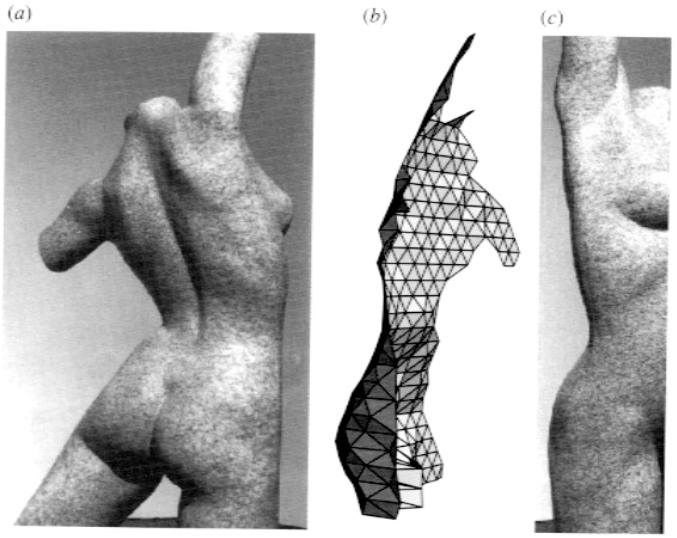

Figure 3. (*a*) The stimulus picture. (*c*) A photograph of the same three-dimensional object, from the same camera position, rotated about the vertical by 90°. The horizontal dimension in this picture is the 'range', that is the objective correlate of the 'depth'. (*b*) A profile rendering of the pictorial relief. One may immediately compare the contours in pictorial and regular space and conclude that the relief is not veridical, although it bears a family resemblance to the depicted object.

of (signed) depth differences should be zero for a consistent surface. (Or, in mathematical terms, one has only a gradient field if the curl vanishes.) In practice I find that, although the total depth difference over the faces is in the ±60 pixels range, the violations are typically in the *subpixel* range. This is of considerable interest in itself, for instance, it falsifies the $2\frac{1}{2}$-dimensional sketch theory (Marr 1982) at the level of consciousness. A single (*local*) setting must involve *global* structure.

Because the data are typically consistent, one may integrate and obtain the depth map of the pictorial surface up to a constant offset (figure 2). This is a convenient representation for the purpose of presentation: one may render the surface in various ways, for example, as a profile view, thus immediately showing the modulations in the depth domain. In cases where one has prepared the pictures oneself one could also photograph the three-dimensional object after a 90° turn about the vertical. Thus one can immediately compare the profile in the pictorial relief with that in regular space. I find that, although there typically is a family resemblance, these profiles differ in quantitative detail (figure 3). Thus pictorial relief is not (significantly) veridical. Moreover, pictorial reliefs of different observers differ. However, the difference between observers is typically limited to a depth scaling (factors up to two are nothing special) and after normalization they are very close (figure 4). This is a finding that fits very well in Hildebrand's (1945) theory of relief perception of sculpture. Pictorial reliefs of different observers *are more like each other than like the real thing*. This might indicate that humans apply the same algorithms. However, one has to be careful here, since a picture can be explained by any member of an

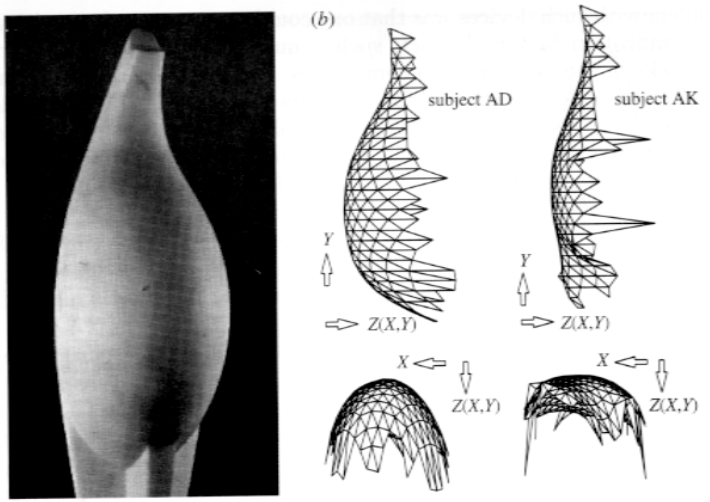

Figure 4. (*a*) The stimulus picture: a photograph of a piece of sculpture by Brancusi (the 'Bird' in the Philadelphia art museum). (*b*) The pictorial relief for two observers AD and AK. Notice that the difference is largely (but not quite) a depth scaling. In this case the depth ratio is about a factor of two, which is nothing special.

equivalence class of scenes of which the original is only one member (Koenderink & van Doorn 1997). When one does not know this class (and at present one is at a loss), one cannot really assess veridicality.

(*b*) *The influence of viewing conditions on pictorial relief*

Since Renaissance times (da Vinci 1989), one has had instructions on how to get the most depth out of a picture: stand in front (oblique viewing flattens) and close one eye. Both instructions make sense, since they reduce the cues that reveal to the observer that the picture is actually flat. It works in the opposite way for a real scene: closing an eye flattens it. Such insights led to the development of effective viewing boxes (von Rohr 1905, 1907, 1920), and at the close of the 19th century to the development of monocular and binocular viewing devices by the optical industry. The *Verant* by Carl Zeiss (designed by von Rohr (1904)) covers one eye and puts the other eye at the perspective centre of the photograph. The centre of rotation of the eye is at the nodal point of a lens system that guarantees a flat field (no accommodation cues) and excellent linearity (eye movements do not reveal inconsistencies). The Verant has been described as 'a stereoscope for single pictures' and indeed, photographs appear in strong pictorial relief, in many respects more satisfactorily so than with a true stereo pair. The *synopter* was another device from the Zeiss factory (von Rohr 1907; Zeiss 1907), this time designed for binocular viewing. The device effectively puts both eyes in coincidence at the midpoint of the binocular segment. A portable version was marketed for use in art galleries. Realistic paintings appear in strong pictorial relief when viewed through the synopter. Turning the synopter on a real scene apparently flattens it.

A problem with such devices was that one could not engrave them with a pictorial relief 'magnification factor'. In fact, such a number would not be determined by the optics alone, the observer also contributes. For instance, one would not expect these devices to work as effectively for observers that are effective monocular (strong eye dominance) or stereo blind. This is probably a major reason that such devices eventually disappeared from the Zeiss catalogue. (The Verant lives on, regrettably in a watered down version, in the common single slide viewer.)

I have quantified the effect of viewing conditions on pictorial relief for a number of (normal) observers (Koenderink *et al.* 1994). I find that viewing a photograph obliquely flattens pictorial relief without affecting the shape, that is to say, the effect is a pure depth scaling. The flattening is a smooth function of viewing angle and can reach values up to a factor of 3–4.

Since an original Zeiss synopter could not be acquired I constructed one by optically cementing together a beam-splitter cube and five 45–90° prisms of the same edge length (figure 5). The device has a fixed interocular distance and about a 25° field of view. Viewing a real scene through the device almost completely flattens it for some observers, it has a strong flattening effect for most. Viewing a picture evokes the (very striking) 'paradoxical stereoscopy' described (but effectively ignored) in the 19th century. The effect is different from observer to observer, in the case of the author monocular pictorial relief is about twice as strong as binocular pictorial relief (on a single picture!) whereas synoptical relief is again about twice that of monocular viewing. Thus the synopter yielded about a fourfold increase of relief compared to binocular viewing. The effect is a pure depth scaling. This is of course to be expected, since the monocular cues that determine the relief are not affected by the viewing conditions at all.

(*c*) *The influence of image structure on pictorial relief*

Image structure (the picture as a 'simultaneous order of pigments distributed over a planar surface') is the 'carrier' of the monocular cues. Whether a particular structure is to be considered a cue for an observer depends both on the picture and on the observer and thus should not be considered a property of the picture. *Cues don't enter the eye*, so to speak. We must assume that observers have internalized certain generic regularities of experience and thus effectively possess models of the causal nexus of the world as it applies to them, one could call these the laws of ecological physics. This may take the form of discursive knowledge (for instance when a landscape painter points out the application of atmospheric perspective to you), but usually observers merely have optical competences on a pre-categorical level. One knows examples of simple neural mechanisms that trigger flight behaviour and implement laws of ecological physics on the hardware level. The well known 'releasers' are probably examples of hard software components (Riedl 1990). One also knows that we can learn to see things without knowing how, many forms of 'expertise' are of that type. In the case of 'experts' the more sophisticated expectations of these observers also play an important role. One often speaks (somewhat nonsensically) of a 'familiarity cue'.

In any case, one expects that changes in image structure might well change the blend of cues for a particular observer, and thus lead to qualitative changes of pictorial relief, quite different from what happens when viewing conditions are manipulated.

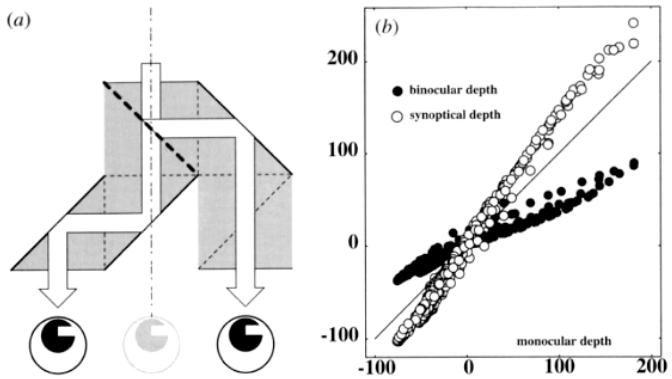

Figure 5. (*a*) A schematic drawing of the synopter device. The beam enters a beam-splitter cube at the top of the drawing. To this cube I have (optically) cemented a pair of 45–90° prisms on the bottom and a triple on the right side (two of this triple merely form a cubical block of glass that equates the optical path lengths for the two split beams). The two eyes (drawn out in black) are optically superimposed in the single 'cyclopean eye' drawn in grey: the user becomes effectively a cyclope. (*b*) Scatterplots of the synoptical and binocular depths (for all vertices) plotted against the monocular depth. (The stimulus was a photograph of a rather eroded piece of Greek sculpture.) Notice that binocular viewing flattens the relief whereas synoptical viewing raises it. The drawn line denotes the identity (monocular depth).

I have systematically varied image structure by photographing the same scene (Koenderink 1996*b*) either by moving the camera or by moving light sources, repainting the objects, etc. Many of such changes can be parametrized in a natural way (for instance, camera and light source through their Cartesian coordinates). Notice that a simple parametric change ('light source 1 m to the left') may lead to numerous changes in the image. (Due to vignetting, interreflection, changes in obliqueness of surface with respect to local net flux vector, etc.) However, this is exactly the desired effect: when one changes local image detail (apparently the simpler method) one almost certainly introduces complicated possible interpretations that are not easily foreseen. Indeed, observers typically fail to notice small parametric variations of camera or light source position which change the whole picture coherently but immediately notice much smaller local changes of image structure (a minor stain say) which are perceived as incoherent with the total scene.

I find that the pictorial relief changes systematically with such parametric variations of the lighting and in *grosso modo* the same way for different observers. The lighting of a scene is particularly interesting because of its many links with art history and artistic practice (Mach 1959; Nurnberg 1948; Clifton 1973; Jacobs 1988; Hunter & Fuqua 1990). When the picture offers only little cues expectations play a major role and observers differ. When the blend of cues is enriched the reliefs become more similar and most observers apparently get the same out of a normal photographic rendering. Although pictorial relief certainly changes systematically with the direction of illumination (say) the changes are modulated on a strong invariant basis: to a large extent 'shape invariance' pertains over large changes of illumination. The remaining

systematic variations reveal idiosyncrasies of the human 'shape from shading (Horn & Brooks 1989)' inferences.

3. Internal representation of pictorial relief

One wonders how pictorial relief is 'represented' in the mind. Most authors silently imply that a depth map would be the most likely candidate. Indeed, depth might be considered a kind of convenient 'common currency' when inferences on the basis of different cues are to be combined. This has been the typical approach in machine vision. The various cues indeed reveal widely different geometrical structure, for instance shading reveals third order curvature structure (Koenderink & van Doorn 1980, 1993*b*), whereas occlusion reveals depth order. It is not known to what extent such data are combined. I think it likely that observers only combine data when *a task requires it* and otherwise simply ignore inconsistencies. This is indeed reasonable since they may take the *consistency of the world* for an axiom. Instead of a single 'internal scene' there may exist only fragments of different natures that may be mined when particular tasks require it, at least when the observer is able to do so. Clearly here lies an enormous field of enquiry waiting to be addressed.

Since the notion of a single internal representation in the form of a depth map that sums up all available structure (conceivably constructed through some kind of Baysian reasoning (Pearl 1988; Riedl 1990), collapsed to a single best guess via some cost function) is, at least implicitly, such a common notion I deemed it necessary to address this issue empirically.

When one assumes that all geometrical structure is derived from a common underlying depth map one concludes that higher order structure (first order: surface attitude, second order: curvature, etc.) will be available at a precision limited by the depth map. If one can show that it is possible to predict depth judgments from attitude judgments but not vice versa one has created a problem for the depth map concept.

In one study (Koenderink 1996*a*) I presented subjects with a picture on which I could superimpose a pair of fiducial points, differentiated ('first' and 'second' for reference) in some manner. The observer is required to judge which of the points is the closer one. I pick the points from the vertices of the underlying triangulation, and always take vertices that belong to a common edge. When the observer has thus probed all edges several times I have obtained the probability of one end being closer than the other for all edges. One can again check consistency by consideration of the faces. I find results that are consistent (within the experimental variation) with an underlying surface (pictorial relief) in all cases. Thus this method yields results not unlike those for the Tissot gauge figure task. Using Thurstone's law of categorical judgment (Torgerson 1958; Coombs 1970) one can again 'integrate' and obtain an explicit depth map. It turns out that such reliefs are similar to those obtained with the Tissot method on the same picture, although different in quantitative detail. Thus one can probe at least two different surface representations in our observer's mind.

I have obtained the probability of one end of an edge of the triangulation being closer than the other directly via the depth difference judgments, but one can also infer it from the Tissot gauge figure data. Thus one can directly compare the empirical spread on repeated measurements for each. I find that the Tissot method yields

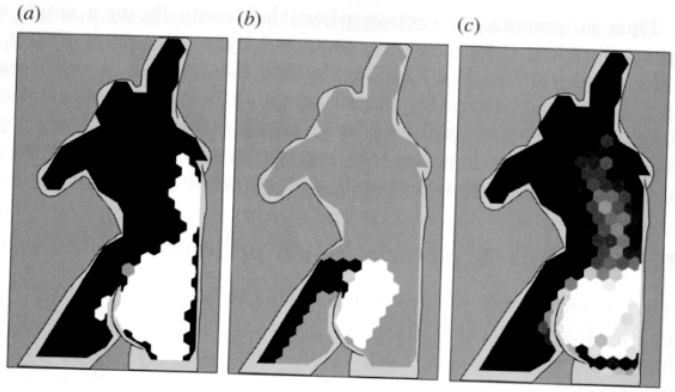

Figure 6. (*a*) The relief obtained with the Tissot gauge figure method thresholded at the dep value of the fiducial vertex. All vertices in front are tinted white, those in the back black. (All vertices that are in front of the fiducial vertex and are connected to it via a path th runs monotonically into depth are tinted white, similarly those in the back are tinted blac (*c*) A map of the probability that the subject will judge a certain vertex to be closer than t fiducial vertex. 'Certainly yonder' is tinted black, 'certainly hither' white, whereas a draw tinted middle grey. Notice that the subject is apparently unable to apply the data structu that one may obtain via the Tissot gauge figure method (*a*).

results that are an order of magnitude more precise than the direct depth differenc judgments. It is thus impossible that the attitude data are derived from the dept data, at least as operationally defined in the ways described.

Of course the results of the method of depth difference judgments might b expected to depend critically on the mutual distance between the fiducial points In the study discussed above this distance (an edge of the triangulation) was equa to the size (length of the major axis) of the Tissot gauge figure. In another study used mutual distances of arbitrary magnitude. Since the combinatorics forbids on to present the observer with all possible pairs of vertices from the triangulation, selected only half a dozen vertices and presented these in combination with all of th other vertices in random order. Thus I obtain maps of the probability of any vertex to be judged closer or more distant than one of the initially selected vertices. Such maps turn out to be particularly revealing (figure 6).

Consider what would be predicted when the observer could simply address the depth map that one can construct on the basis of the Tissot gauge figure measurements: one could threshold at the depth of the fixed vertex and thus split the vertices into a set of closer and more distant ones. Only for a small subset of vertices would the issue be undecided because of the empirical uncertainty. Certainly the precision of judgments would not depend on the mutual distance. As it turns out *this prediction fails miserably*. I must conclude that the observer has no access to a data structure such as the integrated Tissot data at all! Perhaps such a data structure does not exist, or perhaps, when it exists, it cannot be addressed by the depth judgment task. What appears to be the case is the following: A judgment as to the depth order of a pair of points can be reasonably certain when the points can be joined by a path that lies wholly in the pictorial surface and that runs monotonically

into depth. Thus judgments are certain when the points lie on a single slope but are most uncertain when the points are separated by a ridge or course of the relief (Maxwell 1859; Cayley 1870; Todd & Reichl 1989; Koenderink & van Doorn 1993*a*, 1994, 1995). I have found that this model accounts largely for the empirical data although observers have some ability to judge depth order for arbitrary separations. However, the precision of such judgments is an order of magnitude worse than that found for judgments that involve points on a single slope of the relief.

4. The structure of pictorial relief: preliminary conclusions

'Pictorial relief' is an element of consciousness that exists when observers view pictorial renderings such as (flat) photographs or paintings of existent or imaginary scenes. It is evidently due to complicated (largely automatic and subconscious) operations on the image structure and cannot be explained away as a simple application of elementary geometry as with binocular stereopsis. In the latter case the actual image structure is immaterial as long as there is any as is evident from the success of random image structure. That pictorial relief arises at all and is remarkably similar for different observers must be due to the fact that we all evolved and grew up in (generically) the same physical world and that we have internalized the same laws of ecological optics.

Human observers are sufficiently similar that one may exploit this perceptual competence in the design of effective viewing devices of which I have discussed two, the Verant and the synopter. The efficacy of such optical instruments depends as much on the brain of the observer as on the optics of the eye, but this is also the case for other instruments designed to aid vision (blind people with intact eye optics find no use for them). That binoculars can be marked with a magnification that appears to be independent of any observer and the synopter cannot, at first sight seems to put the latter in a corner of instruments that depend on subjective factors. However, observers also have different judgments as to the magnification of binoculars. (For most people binoculars don't 'magnify' at all but merely bring scenes closer and flatten relief, etc.) The differences are really slight although textbooks of optics make a big issue of them. The point is that optical designers simply ignore the user and cheerfully hope for the best.

I have shown that one may perform geometrical measurements in pictorial space much as we would in the space we move in. Geometrical objects revealed (or rather: *defined*) through such measurements cannot be expected to be mutually consistent. For instance, when one measures the variation of surface attitude and the curvature then the latter may well fail to be the spatial derivative of the former. One may not approach pictorial space with the kind of innocence that goes unpunished in regular space. Here lies a large field of endeavour that remains largely unexplored.

The empirical study of the dependence of pictorial relief on image structure reveals the extent and the nature to which observers have internalized laws of ecological optics. Notice that one cannot approach such problems in the same way as binocular stereopsis or visual acuity. For instance, visual acuity depends only on the *internal* structure of the observer, but pictorial relief depends on the generic structure of experiences with the *external* physical world in which optics plays a role.

Many of the concepts that have been lifted by psychology from the machine vision literature may well turn out to be unfruitful (although they may prove of value in the framing of empirical questions). For instance, I have shown that the notion of a depth map as summary representation of pictorial relief is hardly tenable. Indeed, I think it likely that mental structure contains various (perhaps mutually inconsistent) fragments of data structures and that only the execution of particular tasks may perhaps draw on a variety of them and lead to some degree of coordination. Observers are quite content to live with any number of mutually inconsistent fragmentary representations since they can blindly depend on the consistency of the physical world. In fact, any true integration is most likely to take place in the actual interactions with that world rather than in the mind, or perhaps better put, in the manifestation of the mind in such interactions.

This work could not have been carried out without the collaboration of several friends. I mention especially Andrea van Doorn and Astrid Kappers, who are at my laboratory as well as James Todd (Columbus, OH) and Joseph Lappin (Vanderbilt, Nashville, TN).

References

Ames Jr, A. 1925 The illusion of depth from single pictures. *J. Opt. Soc. Am.* **10**, 137–148.

Berkeley, G. B. 1709 *An essay towards a new theory of vision*, 1st edn. Dublin.

Brewster, D. 1844*a*, *b* On the law of visible position in single and binocular vision, and on the representation of solid figures by the union of dissimilar plane pictures on the retina. *Trans. R. Soc. Edinb.* **15**, 349–368, 663–674.

Cayley, A. 1870 On contour and slope lines. *The London, Edinburgh and Dublin Phil. Mag. J. Sci.* **18**, 264–268.

Claparède, E. 1904 Stereoscopie monoculaire paradoxale. *Ann. d'Oculistique* **132**, 465–466.

Clifton, J. 1973 *The eye of the artist.* Westport, CT: North Light.

Coombs, C. H., Dawes, R. M. & Tversky, A. 1970 *Mathematical psychology.* Englewood Cliffs, NJ: Prentice-Hall.

da Vinci, L. 1989 *Leonardo on painting* (ed. M. Kemp and trans. M. Kemp & M. Walker) New Haven, CT: Yale University Press.

Eaton, E. M. 1919 The visual perception of solid form. *Br. J. Ophthalm.* **3**, 349–363, 399–408.

Emerson, E. 1863 On the perception of relief. *Br. J. Photography* **10**, 10–11.

Erens, R. G. F., Kappers, A. M. L. & Koenderink, J. J. 1991 Limits on the perception of local shape from shading. In *Studies in perception and action* (ed. P. J. Beek, R. J. Bootsma & P. C. W. van Wieringen), pp. 65–71. Amsterdam.

Erens, R. G. F., Kappers, A. M. L. & Koenderink, J. J. 1993*a*, *b* Perception of local shape from shading. *Perception Psychophysics* **54**, 145–156, 334–342.

Gibson, J. J. 1950 *The perception of the visual world.* Boston, MA: Houghton Mifflin.

Hildebrand, A. 1945 *The problem of form in painting and sculpture.* (trans. M. Meyer & R. M. Ogden). New York: G. E. Stechert.

Horn, B. K. P. & Brooks, M. J. 1989 *Shape from shading.* Cambridge, MA: The MIT Press.

Hunter, F. & Fuqua, P. 1990 *Light, science and magic, an introduction to photographic lighting.* Boston, MA: Focal Press.

Jacobs, T. S. 1988 *Light for the artist.* New York: Watson-Guptill.

Koenderink, J. J. & van Doorn, A. J. 1980 Photometric invariants related to solid shape. *Optica Acta* **27**, 981–996.

Koenderink, J. J. & van Doorn, A. J. 1993*a* Local features of smooth shapes: ridges and courses. In *Geometric methods in computer vision II* (ed. B. C. Vemuri), vol. 2031, pp. 212–223. Society of Photo-Optical Instrumentation Engineers.

Koenderink, J. J. & van Doorn, A. J. 1993*b* Illuminance critical points on generic smooth surfaces. *J. Opt. Soc. Am.* A **10**, 844–854.

Koenderink, J. J. & van Doorn, A. J. 1994 Two-plus-one-dimensional differential geometry. *Pattern Recognition Lett.* **15**, 439–443.

Koenderink, J. J. & van Doorn, A. J. 1995 Relief: pictorial and otherwise. *Image Vision Computing* **13**, 321–334.

Koenderink, J. J. & van Doorn, A. J. 1997 The generic bilinear calibration–estimation problem. *Int. J. Computer Vision* **23**, 217–234.

Koenderink, J. J., van Doorn, A. J. & Kappers, A. M. L. 1992 Surface perception in pictures. *Perception Psychophysics* **52**, 487–496.

Koenderink, J. J., van Doorn, A. J. & Kappers, A. M. L. 1994 On so-called paradoxical monocular stereoscopy. *Perception* **23**, 583–594.

Koenderink, J. J., van Doorn, A. J. & Kappers, A. M. L. 1996*a* Pictorial surface attitude and local depth comparisons. *Perception Psychophysics* **58**, 163–173.

Koenderink, J. J., van Doorn, A. J., Christou, C. & Lappin, J. S. 1996*b* Shape constancy in pictorial relief. *Perception* **25**, 155–164.

Mach, E. 1959 *The analysis of sensations and the relation of the physical to the psychical* (revised by S. Waterlow). New York: Dover. (Original work published in German in 1886.)

Marr, D. 1982 *Vision*. San Francisco, CA: W. H. Freeman.

Maxwell, J. C. 1859 On hills and dales. *The London, Edinburgh and Dublin Phil. Mag. J. Sci.* **40**, 421–425.

Nurnberg, W. 1948 *Lighting for portraiture*. London: The Focal Press.

Pearl, J. 1988 *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.

Pettigrew, J. D. 1986 The evolution of binocular vision. In *Visual neuroscience* (ed. J. D. Pettigrew, K. J. Sanderson & W. R. Levick), pp. 208–222. London: Cambridge University Press.

Riedl, R. 1990 *Die Ordnung des Lebendigen*. München and Zürich: Piper.

Schlosberg, H. 1941 Stereoscopic depth from single pictures. *Am. J. Psychol.* **54**, 601–605.

Stevens, K. A. 1983 Surface tilt (the direction of slant): a neglected psychophysical variable. *Perception Psychophysics* **33**, 241–250.

Streif, J. 1923 Die binokulare Verflächung von Bildern, ein vielseitig bedeutsames Sehproblem. *Klin. Monatsbl. Augenheilkunde* **70**, 1–17.

Tissot, A. 1887 *Die Netzentwürfe geographischer Karten nebst Aufgaben über Abbildung beliebiger Flächen auf einander von A. Tissot* (autorisierte deutsche Bearbeitung besorgt von E. H. H. Hammer). Stuttgart: Metzler.

Todd, J. T. & Reichel, F. D. 1989 Ordinal structure in the visual perception and cognition of smooth surfaces. *Psych. Rev.* **96**, 643–657.

Torgerson, W. S. 1958 *Theory and methods of scaling*. New York: Wiley.

Truxal, J. G. 1955 *Automatic feedback control system synthesis*. New York: McGraw-Hill.

von Helmholtz, H. 1896 *Handbuch der physiologischen Optik*, 2nd edn. Hamburg and Leipzig: von Leopold Voss.

von Rohr, M. 1904 Linsensystem zum einaugigen Betrachten einer in der Brennebene befindlichen Photographie. Kaiserliches Patentambt, Patentschrift No. 151 312, *Klasse* 42*h*.

von Rohr, M. 1905 Über perspektivische Darstellungen und die Hülfsmittel zu ihrem Verständnis. *Z. Instrumentenkunde* **XXV**, 293–305, 329–339, 361–371.

von Rohr, M. 1907 Über Einrichtungen zur subjektiven Demonstration der verschiedenen Fälle der durch das beidäugige Sehen vermittelden Raumanschauung. *Z. Sinnesphysiol.* **41**, 408–429.

von Rohr, M. 1920 *Die binokularen Instrumente*, 2nd edn. Berlin: Julius Springer.

Wheatstone, C. 1838 Contributions to the physiology of vision. I. On some remarkable and hitherto unobserved phenomena of binocular vision. *Phil. Trans. R. Soc. Lond.* **128**, 371–94.

Wheatstone, C. 1852 Contributions to the physiology of vision. II. On some remarkable and hitherto unobserved phenomena of binocular vision. *Phil. Trans. R. Soc. Lond.* **142**, 1–17.

Zeiss, C. 1907 Instrument zum beidäugigen Betrachten von Gemälden u.dgl., das aus einer geraden Zahl gegen die Mittellinie des Objektraums um 45° geneigter Spiegel in oder außer Verbindung mit einem Fernrohrsystem besteht. Kaiserliches Patentambt, Patentschrift No. 194 480, *Klasse* 42*h*, *Gruppe* 34.

Discussion

H. C. Longuet-Higgins (*Laboratory of Experimental Psychology, University of Sussex, Brighton, UK*). What is the moral of this for computer vision? Should we be making use of any of these assumptions?

J. J. Koenderink. Computer vision will have a hard time doing as well as the human brain. We (humans) use *all* the cues; shading is very important, reflections in the room or in concavities in the object also provide information. It is impossible to input all this information into any algorithm we have today. The systematic results I presented show that different people tend to do the same thing and that there are some human capabilities that computer vision isn't even close to at present.

W. Triggs (*INRIA, France*). In shape-from-shading one tries to get global shape. For humans, getting global shape is hard; we get just local shape in an area and find it hard to determine relative depth, etc. In the light of this, does Professor Koenderink really think we do so much better than computer vision algorithms?

J. J. Koenderink. Yes, I do. Of course, everything is dependent on the information that is actually used, I don't think humans use just shape from shading, they use it in combination with other cues, contours for example. The human observer typically perceives the global shape before the local structure: this requires a professional, analytical attitude. Observers are largely ignorant of the cues they use, they just *see*. The way in which observers, subconsciously, use optical surface properties, interreflections, etc., is way beyond the capabilities of today's computer vision algorithms. Of course, I expect that there will be algorithms in the far future which can outperform humans.

A. Fitzgibbon (*Department of Engineering, University of Oxford, UK*). Do the results perhaps become veridical after some simple transformation of space, for example, a projectivity or affinity?

J. J. Koenderink. No, if we compare the results for different subjects we find that the differences are not global. For example, on the human torso, different parts (e.g. an arm, etc.) may be transformed differently. For particular parts transformations are often approximately rotations in pictorial space, but this is not true globally.

D. Foster (*Aston University, UK*). Further to the questions of failure in veridicality, is there a correlation of successful results with prior knowledge held by the subject? For instance, if we gave the subjects a randomly contoured stone would the errors have a different pattern?

J. J. Koenderink. We have experiments where we compare naive with non-naive subjects and yes, the results from the non-naive subjects are more veridical than