
On so-called paradoxical monocular stereoscopy

Jan J Koenderink, Andrea J van Doorn, Astrid M L Kappers

Utrecht Biophysics Research Institute, Buys Ballot Laboratory, Utrecht University, PO Box 80000, 3508 TA Utrecht, The Netherlands

Received 17 June 1993, in revised form 19 November 1993

Abstract. Human observers are apparently well able to judge properties of 'three-dimensional objects' on the basis of flat pictures such as photographs of physical objects. They obtain this 'pictorial relief' without much conscious effort and with little interference from the (flat) picture surface. Methods for 'magnifying' pictorial relief from single pictures include viewing instructions as well as a variety of monocular and binocular 'viewboxes'. Such devices are reputed to yield highly increased pictorial depth, though no methodologies for the objective verification of such claims exist. A binocular viewbox has been reconstructed and pictorial relief under monocular, 'synoptic', and natural binocular viewing is described. The results corroborate and go beyond early introspective reports and turn out to pose intriguing problems for modern research.

1 Introduction

In daily life human observers routinely judge properties of 'three-dimensional pictorial objects' on the basis of flat pictures such as photographs of physical objects (Gibson 1950) or computer graphics renderings of mathematical objects (Banchoff 1990). This remarkable competence is highly important in many professions and often vital in communication and teaching. It is also a source of wonder and entertainment. One need not be surprised then that efforts have been made to strengthen even further this natural competence: methods for 'magnifying' pictorial relief in single pictures have been recommended at least since renaissance times (Leonardo da Vinci, see Richter 1977). This has given rise to the technology of viewing aids. Viewboxes have been popular for centuries (von Rohr 1905), witness historical prints (Kemp 1990). This technological interest caused a need for increased scientific understanding. Historically, a sharp caesura occurs with the invention of the stereoscope. This invention effectively slowed down the development of monocular viewboxes, because the vivid pictorial relief under monocular viewing was experienced as paradoxical (Claparède 1904) and as perhaps something one might better ignore: any depth not based upon stereopsis was suspect. A new breakthrough in viewbox technology only occurred much later when (at about the turn of the century) Carl Zeiss patented the 'Verant' (von Rohr 1904), based on Gullstrand's theory of vision, with moving eyes for the monocular viewing of photographic prints.

The Verant remains the most advanced viewbox today, and its principles may serve to elucidate the principles of viewboxes in general. (Earlier—and many contemporary—viewboxes use only a subset of these principles.) These are the main points: The viewer is restricted to monocular viewing from a fixed vantage point. The vantage point is placed frontally and centred with respect to the picture. These features agree with the usual recommendations by painters: stand in front of the painting, close one eye, and keep still (Leonardo da Vinci, see Richter 1977). The vantage point is placed at the (effective) centre of perspective, determined by the linear magnification times the focal length of the camera objective from the photographic enlargement. Such conditions can also be obtained with a fixed peephole or sight, though in the Verant a viewing lens is used. The lens places the picture at optical infinity and removes accommodation changes when the eye turns from the

centre (nearest) to a corner (farther away) of the picture. Small translations of the eye do not lead to movement parallax because the picture is at the focal plane of the viewing lens. Such conditions were already realised in many of the early viewboxes, though the Verant has a flatter field and a generally much better off-axis correction than most viewbox optics. Last, the exit pupil of the viewing lens is placed at the centre of rotation of the eye (according to advice by Gullstrand in a letter to Carl Zeiss of 6 July 1902; see von Rohr 1904). This effectively rules out 'keyhole vision'.

A few years after the introduction of the Verant Carl Zeiss also patented devices for the binocular viewing of single pictures (von Rohr 1907; Carl Zeiss 1907). These devices are reputed to yield highly increased pictorial depth, though obviously based on the removal rather than the exploitation of binocular stereo information. This makes perfect sense for single pictures because stereo information for a single picture can only reveal its intrinsic flatness. Removal of any cue that the picture is really that—a (flat) picture—leaves additional leverage to the monocular cues. The most perfect Zeiss design is the synopter (see figure 1), though eventually a cheaper and rather inferior design was actually marketed and apparently sold for the benefit of visitors to art galleries. (The cheaper design used mirrors to place the eyes effectively behind, rather than next to, each other. The resulting size difference is hardly noticed in practice, when viewing distance is over a metre.) The synopter places the eyes effectively at coincident positions, located at the midpoint of the interocular segment. Photographs viewed with the synopter are seen in vivid 'stereoscopic' depth. To many observers the effect is even 'better' than that obtained with regular stereoscopy, because the annoying coulisses effect that so often spoils stereoscopic depth is absent. When real scenes (that is, three-dimensional, not flat, scenes) are viewed with the synopter they take on an uncanny flat appearance of a kind one cannot obtain by just closing one eye. There are several accompanying effects, for example size invariance is much reduced or absent. This is also understandable (at least on a superficial level), for though the brain *should* receive disparity information (because it is looking with both eyes open), it actually receives a vanishing disparity field, indicating a scene at optical infinity. These phenomena are very striking and to judge from the spontaneous introspective reports of (professional) visitors, apparently unexpected to most people.

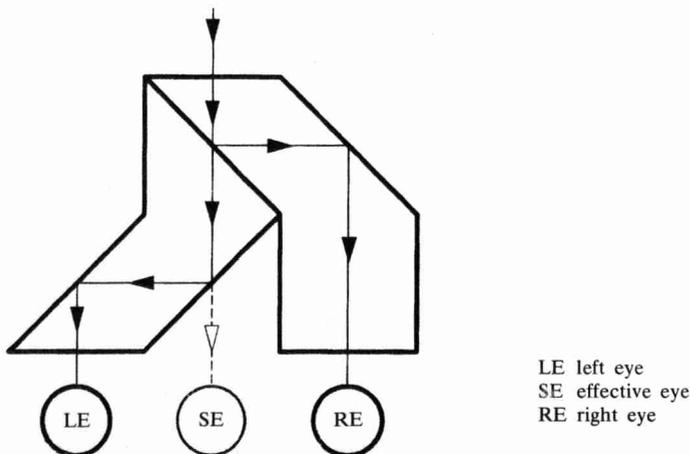


Figure 1. The optical system of the synopter. Only the glass-air and the 50% reflection mirror have been drawn: the blocks of glass have been constructed from five commercially available isosceles rectangular prisms, cemented with Canada balsam. Owing to the antireflection coatings the device probably outclasses the original Zeiss product. Notice that the optical path lengths for the two eyes are equal. The position of the 'effective' eye has been corrected for the refractive index (1.5) of the glass block. Both eyes optically coincide, though depending on the interocular separation the positions may be offset by a few millimetres.

2 Methods

Objective geometrical observations typically imply a judgment of fit of some gauge object to the object to be measured. Yardsticks, dividers, and templates are prototypical examples of gauges. One places a yardstick next to the object and judges optically whether the endpoints of the yardstick coincide with the stretch to be measured. It is often possible to judge the apparent fit of the depicted gauge figure in the depicted scene, though some prior assumptions may be necessary. (In the majority of cases such assumptions are also necessary for the 'real' observation, though this does not seem to worry most people.) We have exploited these simple facts to devise novel ways of gauging aspects of pictorial depth.

In this study we employed gauge figures such as shown in figure 2. The gauge figure was superimposed over the picture at some fiducial location and then was manually controlled via a computer mouse and adjusted by the subject so as to fit the local surface attitude visually—or rather 'pictorially' (Koenderink et al 1992). The nature of the fit is defined via the instruction that the elliptical circumference of the gauge figure should appear as the outline of a circular disc painted on the surface. The 'axle' sticking out at one side effectively removes the 180° ambiguity of the tilt by specifying which side of the circular disc is 'up'. Observers found that this task was easy and natural; typically it took only a few seconds to perform, even without any prior practice. We have not been able to detect effects of practice or training, probably because this task is so simple.

In the actual experiment the gauge figure was presented at the location of a randomly selected vertex of a regular triangulation (we used about 256 vertices) of the interior of the silhouette of the depicted object. The triangulation was, of course, never presented as such to the subject (who saw only the picture and the

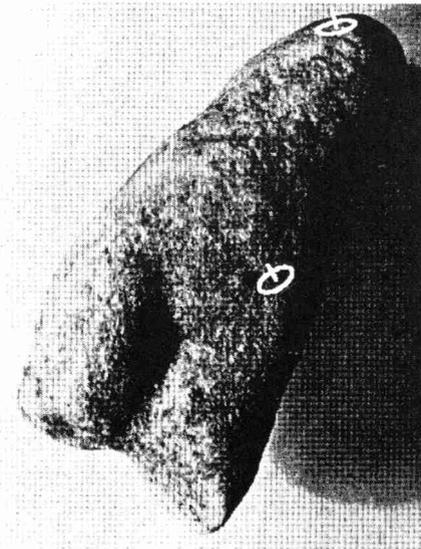


Figure 2. Photograph of a Greek sculpture found in a stream (Hale 1980, page 54). The physical object has both a roundness and a general slant. The picture is scanned to 356 pixels \times 454 pixels and 8-bit grey scale and is displayed in greys on a colour monitor. The gauge figure is displayed in red outline and is the orthographic projection of a disc with an axle sticking out at right angles. (In this figure the gauge figure in the lower part of the picture obviously 'does not fit', whereas the gauge figure in the top part should fit much better visually—according to the reader's viewing conditions.) The picture subtends 12.56 cm by 16.02 cm and is viewed from 75 cm. The gauge figure subtends 32 pixels across.

superimposed gauge figure), but it structured the experiment and enabled us to process the responses in a simple manner after completion of the session.

We compared three different viewing conditions.

- (i) Natural monocular viewing. Only one eye was used for viewing the picture, whereas the other eye was covered by a patch. The vantage point was secured by way of a head-and-chin rest.
- (ii) Natural binocular viewing. Both eyes were used for viewing the picture; the head position (left and right vantage points, or 'cyclopean vantage point', that is, the midpoint between the eyes) was secured by way of a head-and-chin rest.
- (iii) Synoptic (figure 1) binocular viewing. Again, both eyes were used for viewing the picture, but now their optical paths were combined by way of the synopter device. As a result the eyes were optically superimposed at the midpoint of their physical locations and both eyes thus received identical optical inputs. The observer became effectively cyclopean. The synopter was mounted on the same frame that also supported the head-and-chin rest.

The subjects (the authors) were selected for normal stereopsis. AK (female) was emmetropic, AD (female) slightly myopic, corrected to normal (-2 diopters), and JK (male) was presbyopic and corrected to infinity.

We studied both interobserver agreement and individual consistency over time. Settings of the gauge figure were interpreted as surface attitude or, equivalently, local depth gradient. An additional consistency check (applicable to single-observer data) was that the curl should vanish for a true gradient field (Spivak 1975). This mathematical condition expresses the integrability of local attitudes, equivalent to the very existence of a coherent surface. Violation of the condition signifies an 'impossible figure' at an infinitesimal scale: Although a global infinite staircase is at least locally possible, a surface that is at all locations locally like an infinite staircase is globally impossible. We used the Penrose-Moore pseudoinverse (Press et al 1989) to solve the resulting overdetermined—very large—linear system of equations for the nearest coherent surface interpretation of the observations in the least-squares sense. This is merely the obvious, straightforward method. Although this step is not strictly necessary—we might equally well discuss the results in terms of the 'observer' slant and tilt—it is useful because it allows us to view the psychophysical response as a surface in three dimensions. If we speak of the pictorial depth, or relief, for a certain observer and a certain picture, we mean the structure of this surface, which has an intuitive appeal. The response surface is just an alternative way of representing the data. The residual deviations from the best-fitting surface can be due to two causes: the random scatter in repeated settings, and the essential surface inconsistencies (failure of the observed slants and tilts to conform to the depth gradient of any surface at all).

3 Stimuli

The stimuli were monochrome photographs, scanned to 8-bit grey-scale resolution. They were presented in monochrome (grey scale) on a colour monitor attached to a Macintosh Quadra 950 computer. Room and picture frame were clearly visible. The screen was placed at 75 cm from the subject. A head-and-chin rest fixed the head position. A fixation mark was not provided; subjects tended to look at the gauge figure during the actual task.

The gauge figure was overlaid in red outline over the monochrome photograph and was the orthographic projection of a disc with an axle sticking out at right angles. The gauge figure subtended 32 pixels across.

We selected pictures (figures 2 and 3) with a rich gamut of monocular cues. Certainly contour, perspective, various types of texture gradient, shading, interposition, shadow on the supporting structures, etc all play a role. We also selected pictures that show objects with nontrivial and not too familiar shapes, smooth curvature landscapes, and not a great deal of three-dimensional texture (eg hair) on the scale of our gauge figure. We found a rich source in photographs of realistic sculpture of a somewhat generalised or abstracted variety. For this study we selected a photograph of a piece of Greek sculpture found in a stream (figure 2; here, centuries of erosion in the stream did the 'generalising' of the torso) and an impressionist rendering in clay of the head of a veiled lady (figure 3, "Madame X" by Medardo Rosso; here the sculptor has drawn on the veil as an excuse for extreme generalisation).

The torso clearly leans backwards, thus there is a strong overall slant, whereas "Madame X" has been photographed frontally. Intuitively the monocular depth cues are rather stronger for the torso than for "Madame X", though we are at a loss how to quantify this. In the remainder of the paper we will refer to these stimuli as 'the torso' and 'Madame X'.

We are not in possession of the geometrical data of the actual physical objects. This is indeed irrelevant to the present discussion because we are currently not interested in veridicality issues, but only in the influence of viewing conditions and interobserver differences.

4 Results

Though the results obtained with the pictures of the torso and Madame X are in most respects rather alike there are also some interesting differences and we discuss the results for these pictures separately. The observations common to both are discussed with those for the torso stimulus.



Figure 3. "Madame X" by Medardo Rosso (1858-1928) (Hamilton 1967). This wax piece dates from 1892 and is at the Cà Pesaro, Museo Internazionale d'Arte Moderna in Venice. The picture is scanned to 305 pixels \times 481 pixels and measures 10.76 cm by 16.97 cm on the monitor screen.

4.1 *The torso*

The pictorial object—its roundness as well as its general attitude—is highly articulated (figure 4). This picture shows the average over three sessions. Even the raw data of a single session yielded a nice smooth result and were indeed hardly distinguishable from the present average. No further smoothing was done, though the very fact that we fitted the closest surface to the data of course removes the components that do not admit *any* surface to begin with. These components prove to be of the order of the scatter in repeated settings and thus carry no information on the pictorial depth at all. Notice that in this three-dimensional representation two dimensions represent the stimulus domain (the frontoparallel view of the picture), whereas the third, orthogonal coordinate represents the response, that is, the depth, as computed from the gauge figure settings.

The pictorial surfaces for the torso stimulus for the three viewing conditions for a single subject are shown in figure 5. It turns out that the response surfaces agree in considerable detail for all subjects (for example, in figure 6, the Spearman rank-order correlations for all viewing conditions are more than 0.95) and for all viewing conditions up to a depth scaling. Thus the rank correlations of depths are virtually perfect for all pairs of observers or conditions. If viewing conditions are switched (eg an eye is closed or opened), for a single observer the depth dimension expands or contracts like the bellows of an accordion; the depth-order correlations are of the order of 0.99 (figure 7). Apparently, different human observers are able to come up

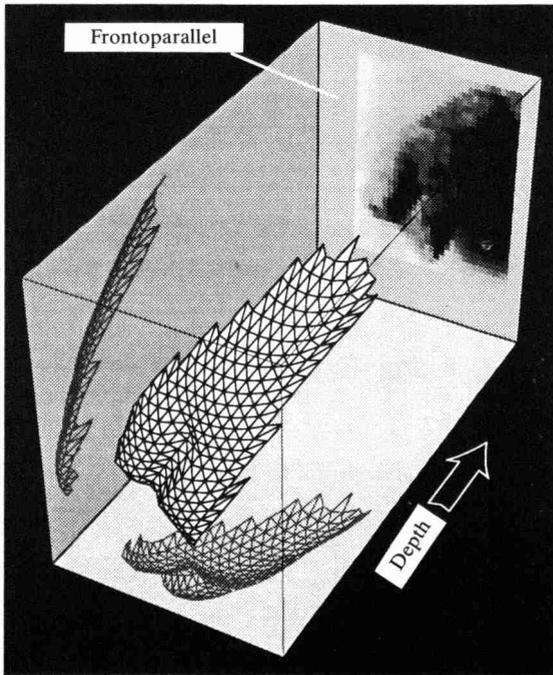


Figure 4. Perspective rendering of the reconstruction (synoptic viewing condition, subject JK, torso stimulus). The picture is shown on the rearmost face of the box, which has the frontoparallel orientation. Drawn are a perspective rendering of the reconstructed pictorial object and its projections on the horizontal and median planes, which extend in the depth direction. Notice the three-dimensional articulation of the pictorial relief, especially visible in the profile views. Notice that one face of the box represents the stimulus domain, whereas the orthogonal direction represents the response continuum. The 'response surface' is technically the graph of the stimulus to response mapping.

with virtually identical and highly articulated three-dimensional interpretations of photographs of objects of nontrivial shapes, even if they have never seen the actual objects. They apparently act on the same objective information concerning spatial extension in the third dimension, because the responses are not merely capricious. This is most remarkable, as even the best modern computer-vision methods are not expected to do as well for this picture (Faugeras, personal communication). Human observers are able to exploit monocular cues to a degree not yet understood, thereby posing a problem for modern visual research.

This perceptual vividness of monocular pictorial relief is rarely mentioned in the current literature but has been well documented and has struck many authors in the past (Ames 1925; Claparède 1904; Eaton 1919; Emerson 1863; Schlosberg 1941); some even speak of 'paradoxical monocular stereoscopy' (Claparède 1904), or a different 'mode of perceiving pictures' (Schlosberg 1941). The paper by Emerson (1863) is historically interesting, since he actually describes the phenomenon but refuses to draw the consequences because these are so obviously scientifically impossible (to him):

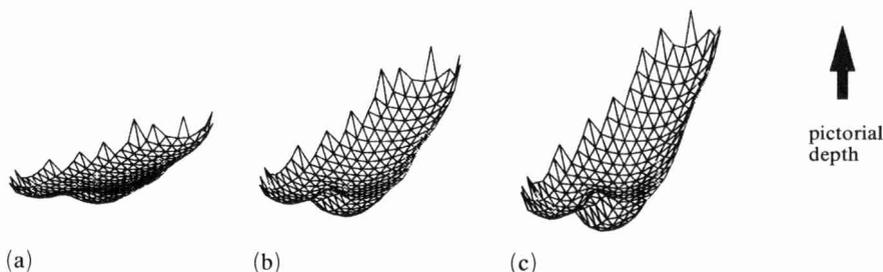


Figure 5. The pictorial surface for (a) binocular, (b) monocular, and (c) synoptic viewing conditions for subject JK and the torso stimulus. The depth is scaled in picture coordinates (pixels), thus the depth range is similar to the frontoparallel size in the synoptic condition. The projection is also shown on the ground plane of the box in figure 4. Synoptic depth is 2.7 times binocular depth. The relation is a perfect linear scaling within the experimental spread.

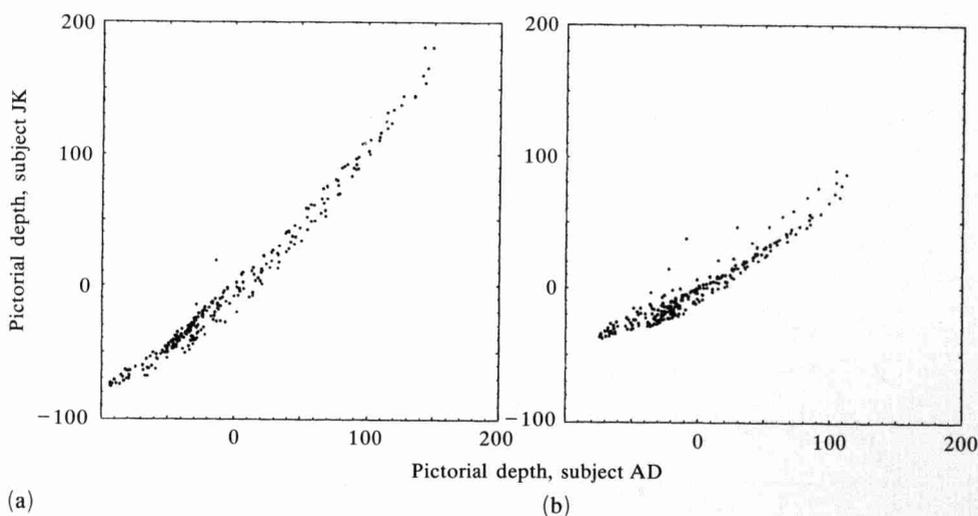


Figure 6. Scatterplots of pictorial depths for subjects AD and JK for the torso picture for (a) monocular and (b) binocular viewing. The Spearman rank correlations are 0.986 for the monocular case and 0.953 for the binocular case; for monocular/binocular per observer: 0.992 for AD and 0.979 for JK. Regression coefficients are 0.977 for monocular data and 1.51 for binocular data. The gain in pictorial relief depth of the monocular over the binocular condition is 1.35 for AD and 2.12 for JK.

he quite literally refuses to believe his eyes. The fact that monocular cues yield highly consistent and detailed spatial perceptions in no way implies conformity of these perceptual judgments to the physical object, of course. In this study we have not even checked veridicality because it is irrelevant to the present issues.

Large interobserver differences exist though individual day-to-day variations are relatively minor. The hypothesis that the observations conform to a consistent surface cannot be rejected in any instance; remarkably, it is as though subjects sample a consistent surface 'in their heads'. Depth-gradient tolerance amounts to about 20%—Weber's Law fits the data very well—in all viewing conditions. Variance in the data is predominantly in the slant, rather than the tilt, dimension. Subjects were reasonably certain of the direction in which the oblique surface tips out of the frontoparallel plane (that is, the tilt of the pictorial surface), but rather less so of the degree by which it tips out of the frontoparallel (that is, the slant). Individual differences and influence of viewing condition are also predominantly to be found in the slant, whereas there is a much greater consensus with regard to the tilt. In summary, all differences are explained as pure depth scalings. (Note that a depth scaling influences only the slant, whereas the tilt is unaffected.)

Depth scalings amount to roughly a factor of 3 for different conditions and a single observer and a factor of 2 for different subjects, thus far exceeding the experimental uncertainties (figures 5–7). Often a perfect fit under one viewing condition or by one subject will be resolutely rejected under another viewing condition or by another subject. The depth scalings for different viewing conditions correlate well with the verbal reports of the subjects, at least in the sense that judgments of greater pictorial depth correlate with wider depth ranges in the observations.

Consider a few numerical data. For subject JK (figure 7) monocular depth exceeds binocular depth by a factor of 2.12, whereas the synoptic adds another factor of 1.39, yielding a total depth gain of about threefold (see figure 5). Depth values for all

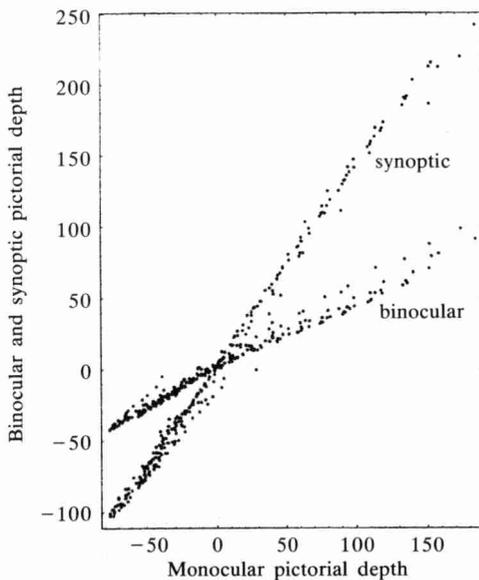


Figure 7. Scatterplots of binocular and synoptic pictorial depths (for the torso picture and subject JK) against monocular pictorial depth. Notice the almost perfect correlation, though the slopes are widely different, indicating a scale change of the pictorial depth domain when viewing conditions are changed. The Spearman rank correlations for the pictorial depths are: monocular–binocular, 0.993; monocular–synoptic, 0.996; binocular–synoptic, 0.987.

conditions are almost perfectly correlated (see figure 7). Such high correlations also apply to intersubject comparisons (see figure 6). For subjects AD and AK synoptic yields hardly any advantage over monocular vision, but gains of 1.43 (AD) and 1.37 (AK) over binocular vision.

The effect for *real* scenes also differs markedly for our observers: for subject JK the synoptic world looked 'flat' (von Rohr 1920), whereas for subject AD it was much like the monocular case. Subject AK reported the same flattening effect for real scenes as did subject JK, though her (objectively measured) pictorial depth seemed to be more in subject AD's category. Here we clearly meet with some puzzling and as yet unresolved issues. It is in principle possible to measure the relief for real objects by methods similar to the one used in this study (by optically projecting an image of the gauge figure onto the physical object), and we will certainly proceed to explore this possibility in the near future. This may enable us to differentiate between observers like AD and AK on the basis of objective data, rather than on the basis of introspection. Records from a dozen visitors reveal that somewhat less than half of them spontaneously reported the synoptic flattening of real scenes, whereas for the remainder synoptic and monocular vision evoked similar impressions. A probably related result has been reported by Richards (1971).

Notice that the effects are apparent both in the overall slant of the torso and in the more local surface articulation (eg the general roundness of the object).

4.2 *Madame X*

The results for the Medardo Rosso piece are in many respects quite similar to those for the Greek torso, though we also find a few puzzling differences. These differences seem to be partly connected with the frontoparallel attitude of the object (the torso is highly slanted as a whole), partly with the rather weaker monocular depth cues presented by the Madame X picture as compared with the torso picture.

Figure 8 shows a comparison of the response surfaces for binocular and monocular viewing in the case of subject AD. Figure 9 shows the comparison between subjects AD and JK. Notice that the response surface (figure 8) is indeed subtly modulated in the depth domain; in the profile view not only gross landmarks such as the nose and the general roundness of the skull, but also articulations due to eye sockets and cheekbones, are apparent. The intersubject correlation (figure 9) is somewhat less than is the case for the torso, although the depth-rank correlations are still high (in every case well over 90%).

For subjects AD and AK synoptic viewing is hardly different from monocular viewing, though both yield increased depth as compared with the binocular viewing condition. For subject JK synoptic viewing typically differs from monocular viewing, though in the case of the Madame X picture the difference is slight, but significant. Thus the synopter yields a less spectacular depth magnification than for the case of the torso picture. The difference is already evident from a comparison of the monocular and the binocular data. From figure 9 one finds that there is hardly a difference between the monocular/binocular gains for subjects AD and JK for the Madame X picture, whereas this difference is appreciable in the case of the torso picture (figure 6).

We conclude that the balance that perception strikes between the flatness of the picture surface and the three-dimensional relief of the pictorial object (due to monocular cues) may shift with the nature of the picture. Apparently the visual system attaches a higher confidence to the pictorial relief (relative to the picture surface) of the torso than to that of Madame X. Intuitively this seems reasonable (many visitors have trouble parsing the Madame X picture at first, whereas the torso evokes an immediate impression of solidity), though we have at present no metric for the strength of the monocular relief information in photographs of natural objects.

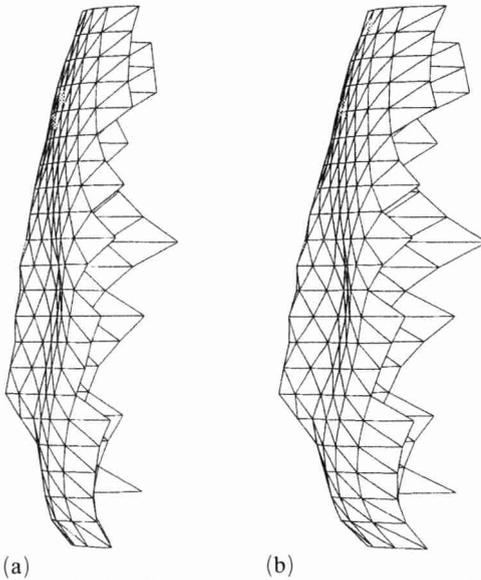


Figure 8. Profile view of the response surface for subject AD and the Madame X stimulus: (a) binocular viewing, (b) monocular viewing. Synoptic viewing is not different from monocular viewing for this subject, though the binocular/monocular conditions show an appreciable depth change. Notice the clear articulation (nose, cheek, eye socket, and roundness and regularity of the face as a whole) despite the rather difficult and blotchy character of the picture.

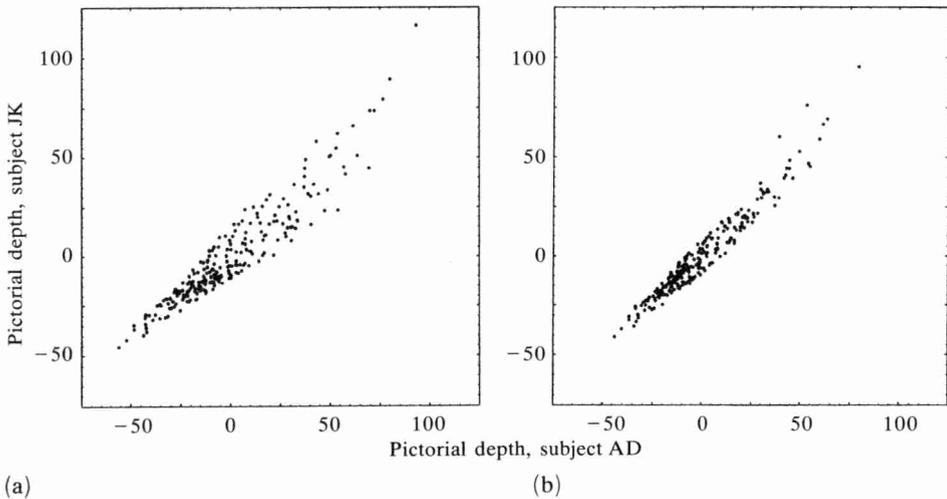


Figure 9. Scatterplots of (a) monocular and (b) binocular pictorial depth for subjects AD and JK for the Madame X picture. Compare these with similar plots for the torso picture, depicted in figure 6. The Madame X results correlate less well than the torso results, a fact that may well derive from the less obvious monocular cues offered by the Madame X picture. The Spearman rank correlations are 0.943 for the monocular case; 0.959 for the binocular case; for monocular/binocular per observer 0.995 for AD and 0.995 for JK. Regression coefficients are 1.04 for the monocular data and 0.973 for the binocular data. The gain in pictorial relief for monocular over binocular viewing is 1.26 for both AD and JK (compare with figure 6 for the torso picture).

A minor but significant effect is that the Pearson rank correlation (a nonparametric method that reveals monotonic functional relations) is higher for the monocular/binocular intraobserver comparison than for the interobserver comparisons, either

monocular or binocular. This indicates that though subjects tend to agree to an astounding degree, they nevertheless differ systematically. This effect is less clear for the torso, and the reason is evident: the torso has a pronounced overall slant that induces a trend in all observers, thus effectively masking interobserver differences.

5 Conclusions

Apparently, human vision involuntarily strikes some kind of compromise between the flatness of the picture surface and the relief due to monocular cues. In binocular viewing the flat picture surface outweighs the relief, whereas in synoptic viewing the monocular relief outweighs the flatness of the picture. The weighing apparently occurs according to some scheme that is reminiscent of Bayesian estimation (Nakayama and Shimojo 1992; Pearl 1988). For all observers with normal stereopsis the closure of one eye yields a large depth gain. For many observers (such as subject JK) synoptic vision offers a dramatic depth gain, for others (such as subjects AD and AK) it is just similar to monocular viewing. The collapse of pictorial relief under binocular viewing occurs for all stereonormal subjects and is often experienced as unexpected, though this apparent flattening of single pictures under binocular viewing is well documented in the older literature (Streif 1923; Tscherning 1900). Tscherning used Javal's 'iconoscope' (which approximates the synoptic viewing condition) and noticed both the flattening of real scenes and the heightened relief (gain over monocular viewing, like subject JK's result) in single pictures. He advanced the same explanation that we propose here.

It is important to appreciate the fact that the present method is just one instance of a large class of possibilities, based upon various gauge figures and instructions. As is indeed true for any observation in the empirical sciences, pictorial depth exists only objectively as an observation. Thus different methods must a priori be expected to yield different and not even necessarily consistent results. For instance, an observation of curvature need not necessarily agree with what can be computed from observed gradients. Although mathematically curvature is nothing but the derivative of the gradient, this by no means implies that observations with different methods are necessarily dependent. If observations of curvature stand in a certain relation to observations of gradient, then this is not (only) a mere mathematical triviality, but an empirical fact that bears on the nature of visual perception. Consequently, comparison of different methods may well be expected to reveal novel aspects of visual functions. Having said this, we hasten to add that this in no way invalidates the present method as a tool for the objective quantification of pictorial depth, or for the assessment of the efficaciousness of viewing aids or of observer competence.

An important finding in this study is that human observers are able to use a variety of monocular information present in photographs of complicated and irregular physical objects in such a way that a highly articulated pictorial object results. Moreover, different observers come up with highly correlated pictorial objects, indicating that we are dealing with some form of objective inference here, not mere capricious fancies. At this moment we are unable to account for such inferences; certainly, present algorithms (shape from contour, from texture gradient, from shading, etc) are not at all in the same league. It is also evident that conflicting disparity information (signaling a flat picture surface) cannot be voluntarily disregarded. Observers do not switch between 'pictorial' and 'object' modes, but somehow manage to reach an uneasy compromise, even though they are well aware that they are attending to a real picture of a virtual physical object. The compromise apparently involves a rather simple type of putting different weights on the various pieces of evidence. Changing the viewing conditions leads to a redistribution of weights, but the subjects also differ in their prior attribution of weights, even though they seem to act on the basis of almost

identical inferences (given the almost perfect rank correlation of depth values). Whereas we are able to understand the shifting of weights for the various viewing conditions qualitatively, we are hard put to understand them quantitatively, nor are we able to understand (or predict from other data) the intersubject differences or the dependence on the image structure. It seems, then, that we are faced with two problems of some scientific interest: first, the construction of the three-dimensional pictorial object from the monocular data, and second the combination of the reconstructed three-dimensional object with the flatness (also reconstructed, but from different data) of the picture surface. Both problems are easily amenable to further scientific investigation.

Acknowledgment. This work was supported by the ESPRIT Basic Research Action 'InSight-II' of the European Commission.

References

- Ames Jr A, 1925 "The illusion of depth from single pictures" *Journal of the Optical Society of America* **10** 137-148
- Banchoff T F, 1990 *Beyond the Third Dimension: Geometry, Computer Graphics, and Higher Dimensions* (New York: Scientific American Library)
- Carl Zeiss Jena, 1907 "Instrument zum beidäugigen Betrachten von Gemälden u.ägl.", Kaiserliches Patentamt, Patentschrift Nr 194480, Klasse 42h, Gruppe 34
- Claparède E, 1904 "Stéréoscopie monoculaire paradoxale" *Annales d'Oculistique* **132** 465-466
- Eaton E M, 1919 "The visual perception of solid form" *British Journal of Ophthalmology* **3** 349-363, 399-408
- Emerson E, 1863 "On the perception of relief" *British Journal of Photography* **10**(181) 10-11
- Gibson J J, 1950 *Perception of the Visual World* (Boston: Houghton Mifflin)
- Hale N C, 1980 *Abstraction in Art and Nature* (New York: Watson-Guption Publications)
- Hamilton G H, 1967 *Painting and Sculpture in Europe 1880-1940* (Harmondsworth, Middx: Penguin Books)
- Kemp M, 1990 *The Science of Art* (New Haven, CT: Yale University Press)
- Koenderink J J, Doorn A J van, Kappers A M L, 1992 "Surface perception in pictures" *Perception & Psychophysics* **52** 487-496
- Nakayama K, Shimojo S, 1992 "Experiencing and perceiving visual surfaces" *Science* **257** 1357-1363
- Pearl J, 1988 *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (San Mateo, CA: Morgan Kaufmann)
- Press W H, Flannery B P, Teukolsky S A, Vetterling W T, 1989 *Numerical Recipes in Pascal* (Cambridge: Cambridge University Press)
- Richards W, 1971 "Size-distance transformations" in *Zeichenerkennung durch biologische und technische Systeme* Eds O-J Grüsser, R Klinke (Berlin: Springer) p 276
- Richter I A (Ed.), 1977 *Selections from the Notebooks of Leonardo da Vinci* (Oxford: Oxford University Press)
- Rohr M von, 1904 "Linsensystem zum einäugigen Betrachten einer in der Brennebene befindlichen Photographie", Kaiserliches Patentamt, Patentschrift Nr 151312, Klasse 42h
- Rohr M von, 1905 "Über perspektivische Darstellungen und die Hilfsmittel zu ihrem Verständnis" *Zeitschrift für Instrumentenkunde* **25** 293-305, 329-339, 361-371
- Rohr M von, 1907 "Über Einrichtungen zur subjektiven Demonstration der verschiedenen Fälle der durch das beidäugige Sehen vermittelten Raumanschauung" *Zeitschrift für Sinnesphysiologie* **41** 408-429
- Rohr M von, 1920 *Die binokularen Instrumente* 2nd edition (Berlin: Springer)
- Schlosberg H, 1941 "Stereoscopic depth from single pictures" *American Journal of Psychology* **54** 601-605
- Spivak M, 1975 *Differential Geometry* volume III (Berkeley, CA: Publish or Perish)
- Streif J, 1923 "Die binokulare Verflächung von Bildern, ein vielseitig bedeutsames Sehproblem" *Klinische Monatsblätter für Augenheilkunde* **70** 1-17
- Tscherning M H E, 1900 *Physiological Optics* (Philadelphia, PA: Keystone Press)