# Binocular depth perception from unpaired image points need not depend on scene organization

Jeffrey J. Tsai
Jonathan D. Victor

Department of Neurology and Neuroscience
Well Medical College of Cornell University
New York, NY 10021

17 text pages
4 Figures

Correspondence: Jeffrey Tsai, Department of Neurology, University of Colorado Health Sciences Center, 4200 E 9th Ave, mail stop b-183, Denver, CO 80262
email: jtsai@med.cornell.edu

Keywords: Occlusion; Binocular depth perception; Stereopsis; Unpaired image points; Stereo correspondence, da Vinci stereopsis

## ABSTRACT

Dichoptic stimuli containing unmatched features can produce depth perception despite the absence of binocular disparity, a phenomenon known as da Vinci stereopsis. Unmatched points can arise from depth discontinuities and partial occlusion in the real world. It has been hypothesized that spatial organization of unmatched image features as dictated by the ecological optics of occlusion might determine perceived depth in da Vinci stereopsis. We tested this hypothesis by creating dichoptic stimuli containing unmatched points in which local cues and overall organization could be dissociated. For these stimuli, observers' perception of depth did not depend on the organization of the scene, but only on the local cues. This finding shows the perceived depth of unpaired points need not depend on reconstructing the spatial organization of depth discontinuities in real-world scenes.

## INTRODUCTION

Binocular disparity is an effective and important cue to relative depth (Howard & Rogers, 1995). Notwithstanding this observation, dichoptic stimuli without apparent disparity signals can produce a sensation of depth (Nakayama & Shimojo, 1990; Anderson, 1994; Liu, 1994; Howard, 1995; Gillam & Nakayama, 1999) not present in either half-image. These stimuli contain elements that lack point-wise correspondence in the two half images, and therefore (by definition) lack disparity. Depth perception based on such unpaired or unmatched features — termed "da Vinci stereopsis" (DVS) (Nakayama & Shimojo, 1990) or "half-occlusion" (Belhumeur & Mumford, 1992) — is qualitatively different from depth perception driven by binocular disparity (Tsai & Victor, 2000) in that discrimination threshold is much greater for DVS than for Wheatstone stereopsis, and perceived depth in DVS is often not veridical. Notably, DVS is distinct from “pictorial” monocular depth cues in that binocular viewing is required for depth perception. Previous studies of binocular depth perception have not provided insight into the mechanism of DVS. On the contrary, most theories of binocular depth perception emphasize the process of establishing binocular correspondence, and view false matches and points that do not have valid matches as targets for suppression, rather than cues to depth (Marr & Poggio, 1976, 1979; Pollard, Mayhew, & Frisby, 1985; Pradzny, 1985; but see McLoughlin & Grossberg, 1998). How, then, is the depth of unmatched image features obtained?

One possibility is that DVS depends on the visual system's interpretation of overall scene organization. Unmatched points in DVS can be related to the presence of partial occlusion in a visual scene. When one surface partially occludes another from view, a wedge of space is visible only to one eye (the eye that partially peeks around the occluder, see Figure 1). That is, an "ecologically plausible" visual scene containing an occluder can generate unpaired features in the resulting retinal images. Note that, as generated by occlusion, unpaired image points are intrinsically ambiguous because they can be localized to an infinite range of depths. Yet they often produce an unambiguous percept at the minimum depth that is compatible with occlusion. An "ecological optics" account of DVS posits that the visual system assigns depth to unpaired points in a manner that reflects this minimum depth consistent with occlusion (Nakayama & Shimojo, 1990; Anderson 1994; Liu et al., 1994;Gillam & Nakayama, 1999).

In this paper, we test the hypothesis that an ecological interpretation of unpaired image points as resulting from occlusion determines the depth perceived in DVS. This hypothesis predicts how the perceived depth in a stimulus will depend on the organization of the visual scene. Stimulus configurations were created in which the locally similar unpaired images required different degrees of depth to be consistent with occlusion in a visual scene. We found that our observers were not sensitive to these manipulations, suggesting that the magnitude of the perceived depth did not reflect the physical depth of the corresponding visual scene.

## METHODS

*Visual Stimulus*

Our stimulus (Fig. 2), based on the “sieve effect” (Howard, 1995), is divided into upper and lower halves, each consisting of 40 "portholes" against a textured background (binary random-checks, 10% black and 90% maximum luminance). A porthole consists of a thin border ("rim") surrounding a region of uniform luminance. The interior of each porthole is assigned one of two luminance values, black or maximum luminance. As in the standard “sieve effect” stimulus, the corresponding portholes in the two half-images have opposite luminance polarity and the rims of all portholes as well as the background have zero disparity. When fused, the sieve effect produces the percept of an uncrossed depth behind the background, seen through the portholes (Howard, 1995; Tsai & Victor, 2000). In these experiments, the portholes are arranged in rows of ten. Within each row, the horizontal positions of the portholes vary randomly, and the vertical positions are similar, subject to a small amount of jitter less than the height of a porthole. In the standard “sieve” stimulus, the luminance assignment of each porthole within one half-image is chosen at random. Here, as we next detail, we use the spatial relationships between adjacent portholes to influence their luminance assignment. This impacts their interpretation in terms of occlusion (see below).

One account of the sieve effect (Howard, 1995) posits that it results from the following visual scene (Fig. 1): two fronto-parallel surfaces arrayed such that an occluding surface (i.e., the textured background) containing a set of apertures (i.e., the portholes) is in front of a surface whose luminance varies across space. Each eye has a limited view of the occluded surface through the portholes. Since the two eyes receive conflicting signals (i.e., unmatched image points) through the portholes, geometry determines the minimum distance ($d_{\mathrm{min}}$) between the two surfaces, similar to the geometric argument for DVS previously noted (Nakayama & Shimojo, 1990). Under the small angle assumption, $d_{\mathrm{min}}$ is equal to the angular subtend of the width of the porthole (Howard & Rogers, 1995; Tsai & Victor, 2000).

This distance is further constrained by global scene organization in the manner described below and illustrated in Fig. 3. Consider two horizontally adjacent portholes separated by a distance *less* than the width of a porthole (Fig. 3a). Geometry requires that the left eye's view through the left porthole must include a portion of the occluded surface that is visible to the right eye through the right porthole. In Figure 3a, this is demonstrated by the left eye's view through porthole *a* (onto region $a_{\mathrm{L}}$ of the rear surface) and the right eye's view through porthole *b* (onto region $b_{\mathrm{R}}$ of the rear surface). For the monocular views to remain consistent with a distance $d_{min}$, the *same* luminance must appear in the left eye's porthole *a* and the right eye's porthole *b*, since these portholes views include common points on the occluded surface. As illustrated, the rule dictated by global scene organization amounts to the following: if two adjacent portholes are separated horizontally by less than a porthole width, they must be of opposite luminance polarity to be consistent with an occluded surface at distance $d_{\mathrm{min}}$. We call this the "no-conflict" condition. On the other hand, if such adjacent portholes have the same luminance, then the stimulus is not consistent with a depth of $d_{min}$ because this would imply the same physical point takes on more than one value ("conflict" condition). The "conflict" condition, however, is consistent with occlusion if the relative distance between the two surfaces is two to three times $d_{\min}$, depending on the horizontal separation of adjacent portholes (again by similar geometric reasoning as above, see Fig. 3b). Thus, by manipulating the luminance relationship of adjacent portholes, we can change the minimum depth that is consistent with occlusion. Our aim then is to determine whether observers could perceive this difference in depth as dictated by occlusion. Note the above rule constraining the luminance seen through adjacent portholes only applies if their horizontal separation is less than a porthole width. For more widely separated portholes, views of the partially occluded surface by each eye are non-overlapping and therefore not subject to conflicts.

Each stimulus contained a "conflict" and a "no-conflict" condition (40 portholes in each) segregated vertically. The "conflict" condition consisted of a variable number of porthole-pairs that were inconsistent with the depth $d_{min}$; the remaining porthole-pairs were consistent with this depth. In the "no-conflict" condition, all porthole-pairs were given luminances consistent with the depth $d_{min}$. Consequent to the rules described above, the "conflict" condition tended to include runs of portholes with the same luminance within a row; while the "no-conflict" condition tended to have alternating luminance values. To minimize the salience of this cue, the fraction of porthole-pairs contributing to the "conflict" cue was kept small (0 to 35%).

Nonius markers were placed along the vertical meridian of the stimulus. Both the textured background and the nonius markers were continuously visible during a trial. At a viewing distance of 114 cm, the entire stimulus subtended 8.3° x 8.3°. Each porthole was 11' on each side surrounded by 2' wide black "rim". Hence $d_{min}$ was 11'.

The half-images were presented via interleaved video frames and polarizing light shutters (Cambridge Research Systems, UK) at a frame rate of 120 Hz. Stimuli were presented using the red gun only because this allowed maximal elimination of leakage through the shutters (Tsai & Victor, 2000). The mean screen luminance was 9.2 cd/m$^2$.

*Subjects and Procedure*

Three subjects with normal visual function who were not aware of the rationale of the experiment participated. Two of the three subjects were experienced psychophysical observers, and had previously participated in a study involving the sieve effect (Tsai & Victor, 2000). In that study, the perceived depth of the sieve effect and its increment threshold were measured by comparison to a disparity depth probe. An uncrossed depth with a finite threshold was found. All three subjects in the current study received practice sessions and were able to see the sieve effect.

Before each trial, the nonius markers were shown on a random-check background. Subjects verified the nonius markers were aligned before initiating the trial. The stimulus was shown for 15 seconds, followed by a random-check mask for 200 ms. The subject could enter a response (and end the trial) at any time after stimulus onset. The subject's task was to decide which depth, as perceived through the portholes in the two stimulus regions (top or bottom), appeared greater. Responses were entered via button-presses without feedback as to their correctness. Subjects were allowed to free-view the stimulus, but nonius markers remained visible throughout stimulus presentation, and subjects were instructed to use them to maintain alignment.

Depth discrimination was measured using a one-interval forced-choice method of constant stimuli. The number of porthole-pairs that were inconsistent with the depth $d_{min}$ varied from 0 to 14. Stimulus conditions were randomly interleaved. For 2 subjects, each data point represented 100 trials. A third subject ran a different paradigm that did not fix the number of trials at each stimulus. For this subject, the total number of trials collected to map the psychometric function was 400.

## RESULTS

Psychometric functions for depth discrimination are shown for three subjects (Fig. 4). The ordinates show the probability of identifying the "conflict" condition as being farther away (in accord with an occlusion interpretation). The abscissas show the number of porthole-pairs in the stimulus whose spatial and luminance relationships are consistent with the "conflict" condition. Error bars indicate 95% confidence intervals (based on a binomial distribution of responses). The larger error bars for FM at the extremes on the abscissa reflect the smaller numbers of trials used for these points. Since our main goal was to determine whether there was any effect of the number of porthole-pairs cueing the "conflict" condition on perceived depth, we fitted the data to a linear function (dotted lines). The y-intercepts are not significantly different from 0.5 for all subjects, indicating that subjects could not discriminate between the "conflict" and "no-conflict" conditions above chance. This is a striking finding because the difference in depth between the two conditions, as predicted by the geometry of occlusion, exceeds the increment threshold for depth perceived in similar stimuli. Specifically, ecological optics predicts a depth difference of 11' to 22', while thresholds for subjects FM and MC are no larger than 6' (Tsai & Victor, 2000). The slopes are not significantly different from zero for all subjects, indicating that performance in the discrimination task is independent of the number of porthole-pairs that cue the "conflict" condition.

One might ask whether the observers' ability to make depth discriminations was somehow impaired by the configuration of the stimulus. For example, since the portholes were interspersed against a zero-disparity background, perhaps the dispersing of disparity signals diminished their discriminability. Another possibility is that the proximity of two groups of portholes cueing different depths ("conflict" vs. "no-conflict") prevents their discrimination. To address this issue, we ran additional tests with one of the subjects (MC), based on Wheatstone stereopsis analogue of these stimuli. In these trials, stimulus configuration differed from the original in that corresponding portholes in the two half images had the *same* luminance but a *non-zero* disparity. In one condition, all portholes had a disparity $d_{min}$ (11'). In the other condition, a proportion of the portholes (equal to the proportion in the "conflict" condition in the main experiment) had a disparity of $2d_{min}$, while the remaining portholes had a disparity $d_{min}$. The subject was asked to judge which of two conditions appeared farther away. MC achieved nearly perfect accuracy in these control trials (one incorrect response out of 600 trials). This indicates that the number and the location of the disparity signals alone cannot explain the poor discriminability in the main experiment. The two stimulus conditions differ in their global scene organization, but are similar in their local characteristics. If the mechanism producing the depth percept depends only on local cues, then the two conditions would generate similar depths that cannot be distinguished from one another, as the data showed.

## DISCUSSION

The advantage of using the sieve effect to study DVS is that there is no binocular matching in the stimulus that would yield a disparity equal to the perceived depth. This property allows for dissociating the role of unmatched points from other processes that generate depth signals, such as depth spreading (Collett, 1985; Buckley et al., 1989), depth interpolation (Mitchison & McKee, 1985, 1987), and double-matching, which can occur when binocular matching is ambiguous. It has been shown that light and dark channels feed independently to the initial stage of binocular matching (Harris & Parker, 1995), and that binocular matching only occurs within a limited interocular contrast ratio (Smallman & McKee, 1995). Therefore, the portholes of the sieve effect, having opposite luminance polarity, are not targets for binocular matching and represent unmatched features. Note that in a standard sieve stimulus, and in the modifications used here, there is no objective depth, and not even an objective depth ordering. Interchanging the left and right half images merely generates another example of the same kind of sieve stimulus, with another (random) assignment of luminances to the portholes.

Depth percepts produced by the sieve effect appear qualitatively different from those produced by binocular disparity. One might wonder whether the percept might be too imprecise for detecting the difference between the "conflict" and the "no-conflict" conditions, or whether subjects perceived the sieve effect at all. In separate experiments, two of the three subjects involved in the current experiment consistently identified an uncrossed depth in the sieve stimulus, as measured by comparison to a traditional disparity probe (Tsai & Victor, 2000). Their depth discrimination threshold was substantially smaller than $d_{min}$ in the current experiment. Thus, the poor discrimination of different stimulus conditions cannot be attributed to the ambiguity of the stimulus alone. Similarly, the control experiment based on a Wheatstone stereopsis analogue of these stimuli shows that the failure of the "conflict" condition to result in a greater perceived depth cannot be attributed to the configuration of the depth cues themselves.

In keeping with an ecological explanation of da Vinci stereopsis (Nakayama & Shimojo, 1990), an interpretation of the sieve stimulus as occlusion dictates a lower bound on the perceived depth. In an earlier study (Tsai & Victor, 2000), the magnitude of the perceived depth reported was sometimes *less* than this lower bound, in violation of the constraints of ecological optics. Although the relationship between neighboring portholes was not controlled in that study, this conclusion remained valid because the presence of a "conflict" configuration would have increased the lower bound. Thus, we can infer that the perceived depth under the "conflict" condition is inconsistent with scene organization to a greater extent than that under the "no conflict" condition. Taken together, these results show that the constraints due to scene organization do not account for perceived depth in the sieve effect.

In summary, we have found the discrimination of perceived depths produced by a stimulus containing unpaired image points, the sieve effect, is independent of manipulations of the stimulus that alter the physical depth as constrained by the ecological optics of occlusion. This finding suggests that the sieve effect is not based on a reconstruction of the overall organization of a visual scene from occlusion cues, but instead depends only on local image cues. More broadly, it implies that ecological optics cannot be the full explanation for da Vinci stereopsis. While we cannot exclude the contribution of a global scene interpretation of occlusion cues for other stimuli, it does not appear to play a measurable role in these stimuli.

## Acknowledgments

This work was supported in part by EY9314 and EY7138 (JDV) and the Cornell-Sloan Kettering-Rockefeller Tri-Institutional MD-PhD Program and GM7739 (JJT). A portion of this material was presented at the Association for Research in Vision and Ophthalmology, Ft. Lauderdale, FL, 2000.

LEGENDS

*Figure 1*. Ecological optics constrains relative depth in a visual scene (bird's eye view). The rear surface is partially visible through apertures (portholes) in the occluding surface. Since its appearance differs as seen by the two eyes (two different regions), the rear surface must be located no less than a distance ($d_{min}$) behind the occluding surface. For small apertures, this minimal distance has a disparity equal to the angular measure of the aperture.

*Figure 2*. An example stimulus. These images have the same relative scale, but differ in absolute dimensions and color, from the actual stimulus. When fused, the stimulus produces the percept of a depth that is behind the background through the portholes. Top half of stimulus represents the "no-conflict" condition. Bottom half contains eight porthole-pairs that are consistent with the "conflict" condition. In this example, occlusion predicts that the depth perceived in the bottom half is greater than that in the top half.

*Figure 3*. Relationships of adjacent portholes constrain relative depth. (a) For an occluded surface at the depth $d_{\text{min}}$, the view through adjacent portholes separated by a distance less than the width of an aperture must have opposite luminance polarity because of the construction of the stimulus and the fact that the same physical location is visible through two portholes. If this rule is satisfied, there exists a self-consistent scene corresponding to the stimulus. (b) If adjacent

portholes have the same luminance, the depth of the plane seen cannot be $d_{min}$ (indicated by cross-hatches), but rather must be least twice as great.

*Figure 4.* Psychometric functions for three subjects in the depth discrimination experiment. The probability of identifying the "conflict" condition as being farther than the "no-conflict" condition is plotted against the number of porthole-pairs consistent with the "conflict" condition. Error bars show 95% confidence intervals. Each data point is the mean of 100 trials for MC and MCr. Confidence intervals are larger for FM at the extremes of abscissa because of a smaller number of trials. The 95% confidence intervals (2-tailed test) of the slope of the best fit line are: MCr (-0.027, 0.030), MC (-0.033, 0.019), FM (-0.022, 0.021). The 95% confidence intervals of the y-intercept are: MCr (0.28, 0.64), MC (0.30, 0.70), FM (0.31, 0.61).

## REFERENCES

Anderson, B. L. (1994). The role of partial occlusion in stereopsis. *Nature, 367,* 365-368.

Belhumeur, P. N., & Mumford, D. (1992). A Bayesian treatment of the stereo correspondence problem using half-occluded regions. *IEEE Proc. Conf. Computer Vision and Pattern Recognition*, 506-512.

Buckley, D., Frisby, J. P., & Mayhew, J. E. (1989). Integration of stereo and texture cues in the formation of discontinuities during three-dimensional surface interpolation. *Perception, 18*, 563-88.

Collett, T. S. (1985). Extrapolating and interpolating surfaces in depth. *Proc R Soc Lond B Biol Sci, 224*, 43-56.

Gillam, B., & Nakayama, K. (1999). Quantitative depth for a phantomsurface can be based on cyclopean occlusion cues alone. *Vision Research, 39*, 109-112.

Harris, J. M., & Parker, A. J. (1995). Independent neural mechanisms for bright and dark information in binocular stereopsis. *Nature, 374*, 808-11.

Howard, I. P. (1995). Depth from binocular rivalry without spatial disparity. *Perception, 24*, 67-74.

Howard, I. P., & Rogers, B. (1995) Binocular vision and stereopsis. New York: Oxford University Press.

Liu, L., Stevenson, S. B., & Schor, C. W. (1994). Quantitative stereoscopic depth without binocular correspondence. *Nature, 367,* 66-68.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194*, 283-7.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proc R Soc Lond B Biol Sci, 204*, 301-28.

McLoughlin, N. P., & Grossberg, S. (1998). Cortical computation of stereo disparity. *Vision Research, 38*, 91-99.

Mitchison, G. J., & McKee, S. P. (1985). Interpolation in stereoscopic matching. *Nature, 315*, 402-4.

Mitchison, G. J., & McKee, S. P. (1987). The resolution of ambiguous stereoscopic matches by interpolation. *Vision Research, 27*, 285-94.

Nakayama, K., & Shimojo, S. (1990). da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Research, 30*, 1811-1825.

Pollard, S. B., Mayhew, J. E. W., & Frisby, J. P. (1985). PMF - a stereo correspondence algorithm using a disparity gradient limit. *Perception, 14*, 449-470.

Prazdny, K. (1985). Detection of binocular disparities. *Biological Cybernetics, 52*, 93-99.

Smallman, H. S., & P., M. S. (1995). A contrast ratio constraint on stereo matching. *Proc. R. Soc. Lond. B, 260*, 265-271.

Tsai, J. J., & Victor, J. D. (2000). Neither occlusion constraint nor binocular disparity accounts for the perceived depth in the 'sieve effect'. *Vision Research, 40*, 2265-76.