Perception & Psychophysics
1999, 61 (1), 144-150

The effects of stereoscopic depth on completion

HIROSHIGE TAKEICHI
Institute of Physical and Chemical Research (RIKEN)
Wako, Saitama, Japan

Stereoscopic depth has a critical effect on completion of partially occluded figures. However, it has not strictly been distinguished whether the effect is direct or indirect through alteration of contour segmentation or parsing. Here, I report that stereoscopic depth does not influence completion of partially occluded figures when parsing is unambiguous from motion cues. This is consistent with the present proposal that stereoscopic depth does not have a unique role in completion and that it is one of the cues to contour segmentation or parsing, which in turn influences completion and surface representation, like motion, shape, or transparency.

Occlusion is one of the difficult problems of visual perception. The random arrangement of objects may unpredictably alter their appearance. Nevertheless, our visual system is able to recognize partially occluded objects. To accomplish this, the visual system supplements the missing parts of the contour of the partially occluded objects. This process is called *amodal completion*.

Stereoscopic depth has a significant effect on amodal completion. At the same time, stereoscopic depth also has a significant effect on contour segmentation or parsing (Nakayama, Shimojo, & Silverman, 1989). Since contour segmentation or parsing is a prerequisite or a part of amodal completion of partially occluded figures, it has not been strictly distinguished whether depth by itself has a significant effect on amodal completion or its effect is indirect through parsing (see also Anderson & Julesz, 1995).

For example, Nakayama and his collaborators (Nakayama et al., 1989; see also Braddick, 1988) have demonstrated that we can recognize a human face in a picture when seen through a horizontal grating. According to them, this is because our visual system veridically identifies which of the potential boundary contours actually belong to the face using stereoscopic depth as a cue to segmentation. In this case, the edges of each strip of the picture segments appear at the same depth as the grating and at a different depth from the face. Therefore, they appear extrinsic to the face (see Figure 1a), enabling amodal completion of the face picture underneath the occluding grating. The same segmentation occurs without a binocular disparity, because other depth cues, such as T-junctions, yield the same depth perception. However, when the depth order of the face picture and the grating is reversed by binocular disparity, the picture is seen as if it is "pasted" on the grating (or "cast" on a screen with cutouts), and the recognition becomes more difficult. This is because the edges of the picture segments now appear at the same depth as the face picture. Consequently, they appear intrinsic to the face (see Figure 1b), thereby disabling the veridical segmentation, parsing, and completion processes necessary for face recognition.

In another example, Nakayama et al. (1989) have demonstrated that a C-like figure appears like an O, when an occluder is seen in front of the C at the opening, but not when the occluder is seen behind (see Figure 2a). In another example, they have demonstrated that we see one or three objects depending on the depth (see Figure 2b). Both have been shown as demonstrations showing an effect of stereoscopic depth on contour segmentation or parsing.

However, there is a difference between the demonstration in Figure 2a and that in Figure 2b. In Figure 2a, we see completion only when the occluder is in front. In contrast, we see completion both when the occluder is in front and when it is in back in Figure 2b. The effect of depth in Figure 2b is not on the occurrence of completion but on the mode or the appearance of the completed figure; we see three disks and one window in one case, and one disk and three windows in the other case. The perceived quality is different, but the shape of the completed figures is the same. The cause of this difference between Figures 2a and 2b is as follows. In Figure 2a, the depth of the occluder changes the parsing of the T-junction formed by the C-figure and the occluder. In contrast, in Figure 2b, the depth does not change the parsing. The two contours of each blob always belong to different entities. Therefore, it does not seem the depth itself but the parsing that is critical to the occurrence and the shape of completion.

Naively, it may seem natural that completion of partially occluded figures occurs only if the occluder is seen in front of the occluded figure, because it is physically

This study was supported by the Inamori Foundation of Science. Parts of the experiments were conducted at the Communications Research Laboratory of the Ministry of Posts and Telecommunications of Japan and at the University of Tokyo. The author is grateful to Keiji Tanaka for his comments on Experiment 3 and to Daniel Palomo for his comments on earlier versions of the manuscript. Correspondence should be addressed to H. Takeichi, The Institute of Physical and Chemical Research (RIKEN), Hirosawa 2-1, Wako, Saitama 351-0198 Japan (e-mail: takeichi@riken.go.jp).

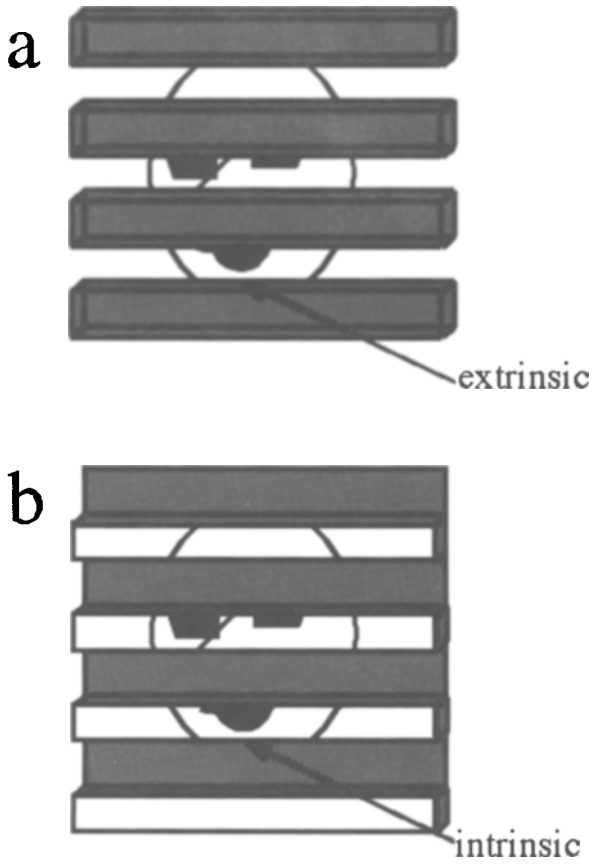

Figure 1. The difference between the face-rearward case and the face-forward case in Nakayama et al. (1989). (a) When the grating is in front of the picture, the border between the picture and the occluding grating is veridically recognized as belonging to the occluder (extrinsic to the picture). Amodal completion underneath the occluder occurs, and the face is easily recognized. (b) When the grating is behind the picture, the border is misrecognized as belonging to the picture (i.e., intrinsic to the picture). Amodal completion is lost, and the recognition becomes difficult.

impossible to "hide" something in front of another. This naive "ecological" reasoning, or "natural intelligence," may not hold if it is not depth itself but parsing that is critical to completion.

The question thus is whether depth has a privileged role in and a direct effect on completion. Shipley and Kellman (1992) have argued that the two types of completion that can be switched by stereoscopic depth—the illusory contour and the amodal completion—are identical, as shown in Figure 2b.

Shape and depth may have a drastic effect on completion. For example, when a revolving disk is seen through a revolving elliptical aperture, as shown in Figure 3, the perception is veridical and the disk appears rigid, although the revolving aperture occludes different parts of the disk to different extents from time to time. This is because the contour segmentation process veridically identifies which of the image contours of the partially occluded disk is intrinsic and which is extrinsic to the disk. The missing part of the disk is completed. In contrast, when the aperture becomes a virtual one by removal of the portion of the aperture frame that does not occlude the disk, the disk no longer appears rigid, and it is seen as a deforming figure as it moves along a square-shaped trajectory. In this case, the intersection of the virtual aperture and the disk forms an L-junction rather than a T-junction. In addition, the previously extrinsic contour of the disk loses its counterpart (the portion of the aperture frame that does not occlude the disk), sharing a common motion. As a result, all the contours of the disk become intrinsic to it, and completion becomes difficult or impossible.

Taken together, it has been suggested that stereoscopic depth does not have a privileged role in amodal completion and that its effect is mainly indirect through an effect on parsing. To test this conjecture, a face-recognition experiment was conducted with an occluder in front and in back, with motion as an unambiguous cue to parsing. More specifically, a face was presented anorthoscopically with a binocular disparity. The face was seen either in front of or behind the occluding aperture frame, depending on the binocular disparity, but the parsing or the segmentation of the face from the aperture frame was always the same due to the motion of the face. If depth directly influences completion, the anorthoscopic perception of the face should be impaired when the face is seen in front of the aperture. In contrast, if depth indirectly influences completion through parsing, it should remain unaffected, because the parsing is the same irrespective of the depth.

Figure 2. (a) A stereogram showing an effect of depth on completion. The diverging fuser should fuse the left and middle images to see the occluder-in-front case and should fuse the middle and right images to see the occluder-in-rear case. The converging fuser should fuse the opposite pairs. The C-like figure is completed as an O-like figure only when the occluder is seen in front. (b) A stereogram showing an effect of depth on completion. The observer is expected to see either a disk and three windows or a window and three disks, depending on the depth. The diverging fuser should fuse the left and middle images to see the one-window–three-disks case and should fuse the middle and right images to see the one-disk–three-windows case. The converging fuser should fuse the opposite pairs.

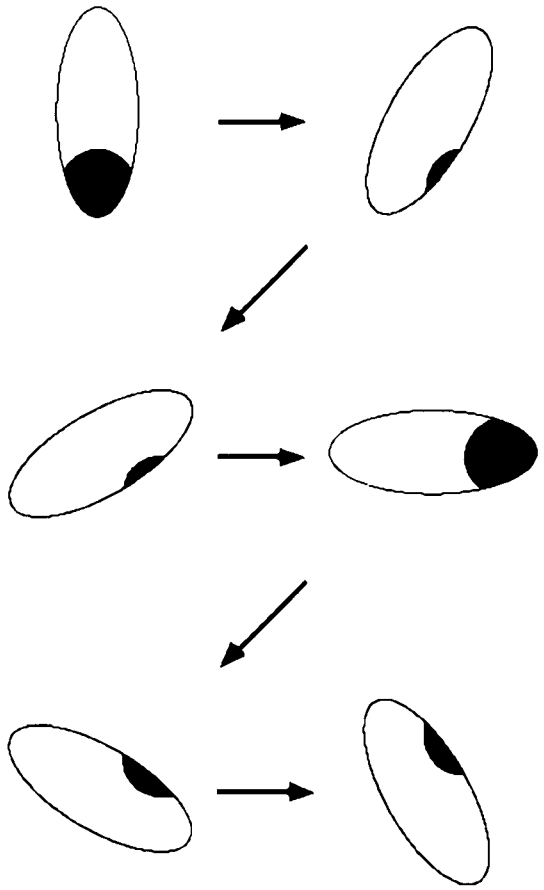

Figure 3. A variation of anorthoscopic perception showing the effect of explicit recognition of the aperture shape on the apparent shape of the occluded figure. A revolving disk is seen through a revolving elliptical aperture. The disk and the aperture revolve in opposing directions. When the elliptical aperture is visible, the disk veridically appears rigid. However, when the elliptical aperture is invisible, the veridical segmentation is lost. The disk appears nonrigid, and it no longer looks like a disk. Note that the motion information is the same in both cases.

EXPERIMENT 1

Method

Subjects. Four naive adults and the author served as subjects. All had normal or corrected-to-normal vision.

Stimulus and Apparatus. One hundred and twenty digitized photographs of human faces were used. They looked similar to each other, because the photographs were taken of people of similar ages (about 22 years old), the same sex (female), the same race (Asian), and without glasses. The photographs were novel to all the subjects except for the author (A) and S1, who had become familiar with them during the preparation of the experiment. Each face was 5.7° of arc high and 4.3° of arc wide. The background of the photograph was erased manually using graphics software.

The stimulus was translated upwards twice past an aperture for each trial. The aperture was embedded in a background pattern of random dots and measured 0.57° of arc high by 7° of arc wide. Depth was defined by a near or far disparity of 17 arc min between the background and the face. The motion was simulated in 92 frames. The apparent velocity was 3.7°/sec. The stimuli were presented on a Mitsubishi JUM-1481A monitor connected to a Commodore Amiga microcomputer at a frame rate of 60 Hz. A phase haploscope (Haitex X-Specs 3D) was used for stereoscopic presentation. With this setup, each eye's image was updated at 30 Hz.

Procedure. The tolerance of face recognition to noise was measured for each of the depth conditions. The noise consisted of the addition of random dots to the stimulus. The noise level was defined as the proportion of the area covered by the dots.

For 3 of the subjects (S1, S2, and A), the performance was measured in terms of the correct response rate. In this procedure, the noise level in each trial was one of 0%, 5%, 10%, 15%, or 20%. There were 12 trials for each of the noise levels and for each of the depth conditions. Each trial was initiated by the subject's pressing a mouse button. First, the sample stimulus was shown for 1 sec. Next, after a 1-sec interval, during which a random dot pattern was shown, either the target or a distractor face was shown anorthoscopically, as described in the Stimulus section. Finally, the subject judged whether the sample and the target were images of the same person. The next trial proceeded immediately, following the subject's response. To discourage a simple template matching strategy, the outer contour of the hair was trimmed to make the apparent shape uniform across different faces, and the mirror image (left–right reversal) of the sample was used as the target.

For the other 2 subjects, a different procedure was employed. Here, threshold noise level was measured by a modified random staircase method. In this method, trials alternated randomly between two series, each of which corresponded to one of the depth conditions. Starting from a level of 0%, the noise was either increased by 2% after two consecutive correct responses or decreased by 2% after one incorrect response on the next confrontation with that series. When the subject made an error at the 0% level, no adjustment was made. However, this did not happen except in the initial stages of the experiment. When the subject made two consecutive correct responses after a wrong response or a wrong response after two consecutive correct responses within a series (a turn), the noise level at that trial was recorded. Data were first collected until the 7th turn for both of the series and then until the difference between the minimum and the maximum noise levels during the last 4 turns became smaller than 8%. The resulting numbers of turns were 8 and 9 for Subject S3 and 9 and 10 for Subject S4. The other aspects were the same as the first procedure except that the presentation of the sample lasted only 670 msec.

The experiment was conducted in a dark room without dark adaptation. There was no fixation point. A chinrest was used in the staircase procedure (Subjects S3 and S4), but not in the first procedure (Subjects S1, S2, and A). The observation distance was about 60 cm.

Results and Discussion

The correct response rate is plotted as a function of the noise level for each subject in Figure 4. An analysis of variance (ANOVA) after arcsine transformation showed a significant effect of noise [$F(4,8) = 7.35, p < .05$]. The effect of depth [$F(1,2) = 0.0610$] and its interaction with noise [$F(4,8) = 0.765$] were not significant ($p > .05$).[1] The overall correct response rates in the face-rearward condition were 91% (A), 91% (S1), and 70% (S2); those in the face-forward condition were 91% (A), 80% (S1), and 75% (S2). The pooled data also did not show a difference [$\chi^2(1) = 0.32, p > .05$]. The mean noise levels for the last 4 turns were 10% (S3) and 17% (S4) for both depth conditions. Thus, Experiment 1 did not show an effect of stereoscopic depth on the anorthoscopic face recognition, suggesting that the effect of stereoscopic

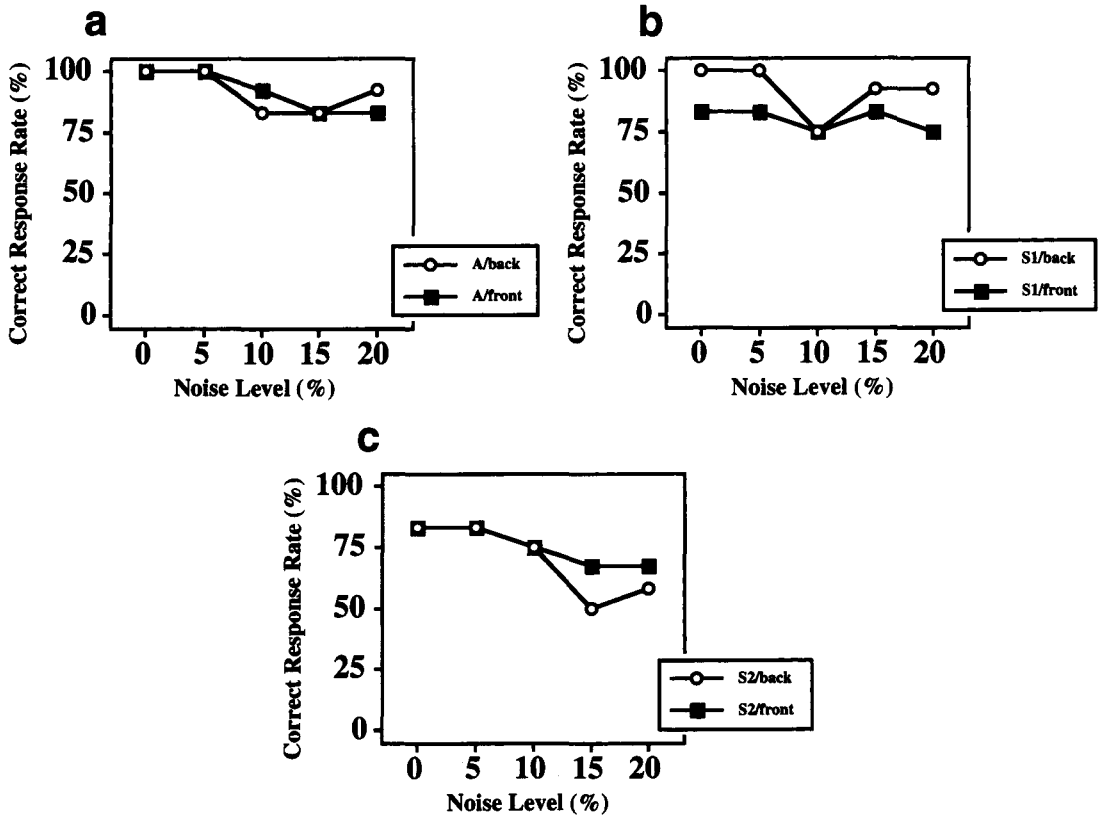

Figure 4. Results of Experiment 1. The correct response rate for each subject is plotted as a function of the noise level. No effect of depth on anorthoscopic shape perception was found.

depth on completion is indirect through parsing, but not direct.[2]

However, it was possible that this negative result was due to inefficiency of the depth cue. To exclude this possibility, the following experiments were conducted. In Experiment 2, Nakayama et al.'s (1989) original experiment was replicated with a setting similar to that of Experiment 1.

EXPERIMENT 2

Method

Subjects. The author (A) and 2 of the previous subjects (S1 and S2) participated in this experiment.

Stimulus and Apparatus. The same stimulus and apparatus as in Experiment 1 were used, except that a horizontal grating was laid over the face pictures in the second presentation. The height of each bar of the grating was 0.7° of arc, and the face picture was divided into six horizontal strips. The face picture was stationary and was presented for 1 sec.

Procedure. Only the first procedure was used. The other aspects were the same as the those of Experiment 1.

Results and Discussion

The correct response rate is plotted as a function of the noise level for each subject in Figure 5. An ANOVA after arcsine transformation did not show a significant effect of noise [$F(4,8) = 3.57, p > .05$]. However, the main effect of depth [$F(1,2) = 127, p < .05$] was significant, and the interaction between depth and noise [$F(4,8) = 0.104$] was not significant ($p > .05$).[3] In addition, the correct recognition rates calculated only from the *same* trials, which are more appropriate for the direct comparison with the original study, were 100% (A), 90% (S1), and 70% (S2) in the face-rearward condition and 93% (A), 73% (S1), and 46% (S2) in the face-forward condition. The pooled data showed a statistically significant difference [$\chi^2(1) = 6.53, p < .05$].

Although the result was in agreement with Nakayama et al.'s (1989) original finding, suggesting that the experimental setup should have been able to reveal potential effects of depth, this is indirect evidence. Thus, in Experiment 3, a dual-task paradigm was used to ensure that the disparity was an effective cue to depth, and a more simplified stimulus was used to avoid possible influence of cognitive factors, such as familiarity.

EXPERIMENT 3

Method

Subjects. Four naive adults and the author served as subjects. None of the naive subjects participated in the previous experiments. All had normal or corrected-to-normal vision.

Stimulus and Apparatus. Twelve dot patterns were used. Each of them contained eight gray dots or blobs (16 cd/m^2). Dots were

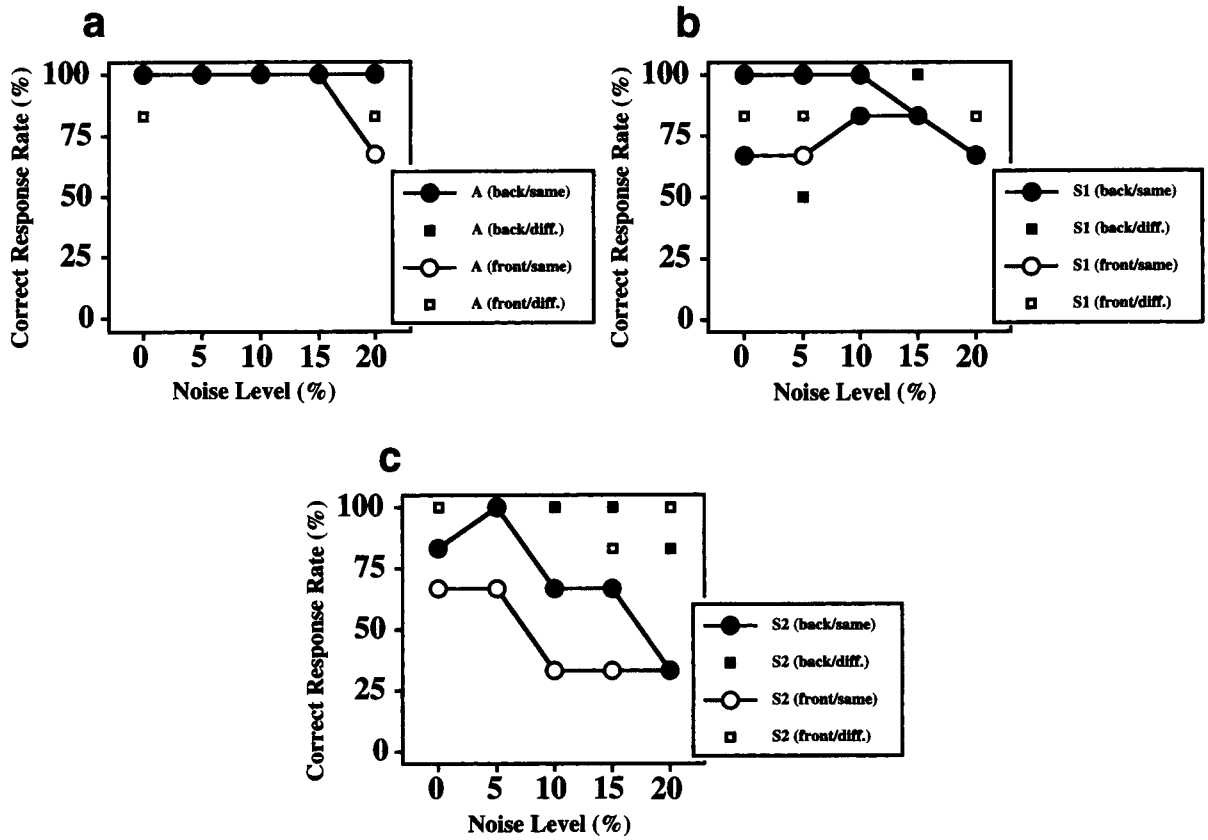

Figure 5. Results of Experiment 2. The correct response rate for each subject for each response type (*same* or *different*) is plotted as a function of the noise level. An effect of depth on static completion was found.

placed on the grid points of a virtual square matrix of four rows × four columns. One arrangement of dots was randomly created, and each of the remaining 11 arrangements was made from the first by randomly choosing and relocating one of the dots. Each pattern differed from the rest. The virtual matrix was 60 arc min in height and width. The dots were 11 arc min in radius. One example is shown in Figure 6. A new set was created at each session.

In each trial, the stimulus was translated past an aperture. Depth was defined in terms of binocular disparity. In the temporal disparity condition, an initially black, vertically oriented aperture (60 arc min height, 9 arc min width) was presented on a gray (30 cd/m^2) background. In the spatial disparity condition, it was horizontally oriented (9 arc min height, 60 arc min width). The aperture was accompanied by two nonius markers, one above and one below.

The stimulus carried either a near disparity or a far disparity relative to the aperture. In the temporal disparity condition, the stimulus was translated horizontally past the vertically oriented aperture. When a figure translates behind an aperture horizontally, the eyes see the same part of the figure at different times. This time difference, which depends on the depth separation between the aperture and the figure, is called *temporal disparity*. I used temporal disparities of −110, −70, 0, 70, and 110 msec. Positive values indicate the precedence of the left eye stimulus in the rightward movement or the right eye stimulus in the leftward movement. Since the direction of the movement and the direction of binocular disparity were the same (i.e. horizontal), these temporal disparities corresponded to spatial disparities of −9, −6, 0, 6, and 9 arc min, where positive values indicate near disparity.

In the spatial disparity condition, the stimulus translated vertically past the horizontally-oriented aperture. I used spatial disparities of −10, −6, 0, 6, or 10 arc min. All the stimuli were presented on a monitor (NEC PC-KD881) connected to a microcomputer (NEC PC-9801NL) at a frame rate of 56 Hz. A mirror haploscope was used for stereoscopic presentation.

Procedure. The experiment was conducted in a dark room without dark adaptation. There was no fixation point. A chinrest was used to maintain an observation distance of 140 cm, where 1 pixel subtended 1 min.

In each trial, the stimulus translated past the aperture three times. A tone was given about 250 msec prior to each presentation. Translation was simulated by successively presenting 20 slices of the pattern with a 36-msec interslice onset asynchrony. In other words, each eye's image was updated at 28 Hz. There was no interslice interval. Thus, the apparent translation velocity was 1.4°/sec. The overall translation duration was 710 msec. The interpresentation interval was about 250 msec. During this period, an empty black aperture was presented.

In the following response phase, the aperture disappeared, and four stimuli were presented in two rows and two columns. The stimuli were entirely and binocularly visible. One of the four stimuli was the one shown in the presentation phase. The remaining three distractors were chosen randomly from the other 11 patterns. The subject was asked to correctly identify the previously seen stimulus from among these four choices using a mouse pointer. At the same time, the subject indicated the direction of depth of the stimulus (vs. its surround) by making this selection with either the left or the right mouse button. The position of the correct stimulus was randomized across trials. Following a response, the next trial proceeded immediately . The subjects were given as long as desired to respond and were free to rest by simply delaying their responses.

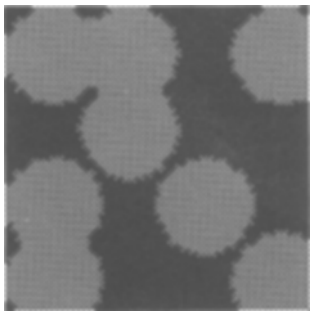

Figure 6. An example of the stimuli used in Experiment 3. Eight gray blobs were arranged in a virtual matrix of 4 × 4.

Each stimulus was presented 10 times to cover all combinations of the two directions by five disparity values within a session. Thus, there were 120 trials in one session. The temporal disparity condition and the spatial disparity condition were conducted in separate sessions.

Results and Discussion

The proportions of correct identification of the depth of the stimulus for the subjects for each level of disparity were 52% (S4, 110 msec or 9 min), 50% (S4, 70 msec or 6 min), 56% (S5, 9 min), 88% (S5, 6 min), 56% (S6, 9 min and 6 min), 62% (S7, 9 min), 81% (S7, 6 min), 77% (A, 9 min) and 83% (A, 6 min) in the temporal disparity condition and 54% (S4, 10 min), 52% (S4, 6 min), 75% (S5, 10 min), 69% (S5, 6 min), 71% (S6, 10 min), 65% (S6, 6 min), 58% (S7, 10 min), 56% (S7, 6 min), 90% (A, 10 min), and 88% (A, 6 min) in the spatial disparity condition, excluding cases of zero disparity. Since the one-tailed binomial test requires more than 65% correct responses for statistical significance at the 5% level, analysis was restricted to the data from Subjects S5, S7, and A (the author) with the disparity of 70 msec (6 min) in the temporal disparity condition and the data from Subjects S5, S6, and A in the spatial disparity condition. Also, only the trials with correct depth identification have been used in the following analyses.

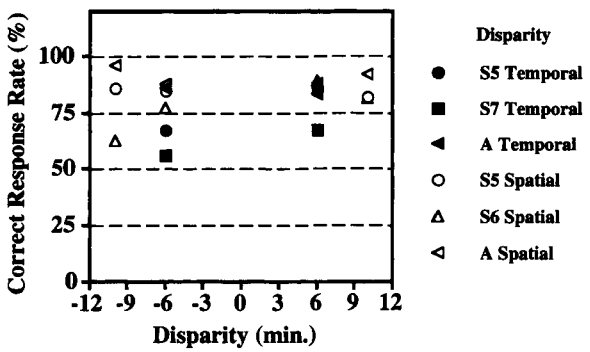

Figure 7. Results of Experiment 3. Correct recognition rate is plotted as a function of disparity. Different symbols represent different subjects. No effect of depth on anorthoscopic shape perception was found. Not all data are included; see text for the detail.

The proportion of correct responses is plotted for each condition, disparity, and subject in Figure 7. The subjects could identify the translating figure significantly better than chance ($p < .05$, one-tailed binomial test) in each of all the disparity conditions. The results indicate that anorthoscopic shape perception remained intact irrespective of binocular disparity and depth perception. The results of an ANOVA after arcsine transformation did not show a significant effect of depth [temporal, $F(1,2) = 4.60$, $p > .05$; spatial, $F(3,6) = 0.0533$, $p > .05$).[4] In fact, the subjects could recognize the stimulus equally well even if it was seen in front of the aperture, supporting the argument that depth does not have a direct effect on completion.

GENERAL DISCUSSION

I have provided data that are in accordance with the idea that depth does not have an effect on completion, when the parsing is unambiguous from a motion cue. This is consistent with the argument that depth does not have a unique role (a "gate") in completion, and it works as one of the cues to contour segmentation or parsing together with motion, shape, or transparency.

One of the major goals of early visual processing is thought to be the formation of surface representation (He & Nakayama, 1994; Nakayama & Shimojo, 1992). Although the exact nature of the biological surface representation has not been clarified, it should include contour segmentation, assignment of "border-belongingness," and completion processes. It is clear that stereoscopic depth plays an important role in the formation of this alleged representation. However, other factors, such as viewpoint invariants (Albert, 1993), local kinematics (Bruno & Gerbino, 1991), transparency (Takeichi, 1998; Trueswell & Hayhoe, 1993), and other forms of natural constraints (Takeichi, Nakazawa, Murakami, & Shimojo, 1995), must also be considered to have the complete understanding of the biological surface representation.

There is another interpretation of the results: It is possible that amodal completion in the static domain and amodal completion in the dynamic domain are different perceptual processes, and, therefore, stereoscopic depth has different effects on different phenomena. Even if this were the case, the present data have shown that (1) stereoscopic depth does not have a significant effect on dynamic completion, (2) completion is not always "gated" by stereoscopic depth, and (3) the simplest explanation of the effect of stereoscopic depth on completion (i.e., an object can occlude another object behind it, but not vice versa) needs to be reconsidered.

REFERENCES

ALBERT, M. K. (1993). Parallelism and the perception of illusory contours. *Perception*, **22**, 589-595.

ANDERSON, B. L., & JULESZ, B. (1995). A theoretical analysis of illusory contour formation in stereopsis. *Psychological Review*, **102**, 705-743.

BRADDICK, O. (1988). Contours revealed by concealment. *Nature*, **333**, 803-804.

BRUNO, N., & GERBINO, W. (1991). Illusory figures based on local kinematics. *Perception*, **20**, 259-274.

HE, Z. J., & NAKAYAMA, K. (1994). Perceiving textures: Beyond filtering. *Vision Research*, **34**, 151-162.

MORGAN, M. J., FINDLAY, J. M., & WATT, R. J. (1982). Aperture viewing: A review and a synthesis. *Quarterly Journal of Experimental Psychology*, **34A**, 211-233.

NAKAYAMA, K., & SHIMOJO, S. (1992). Experiencing and perceiving visual surfaces. *Science*, **257**, 1357-1363.

NAKAYAMA, K., SHIMOJO, S., & SILVERMAN, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, **18**, 55-68.

SHIPLEY, T. F., & KELLMAN, P. J. (1992). Perception of partly occluded objects and illusory figures: Evidence for an identify hypothesis. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 106-120.

TAKEICHI, H. (1998). *Perception of transparency from T-junctions*. Manuscript in preparation.

TAKEICHI, H., NAKAZAWA, H., MURAKAMI, I., & SHIMOJO, S. (1995). The theory of the curvature-constraint line for amodal completion. *Perception*, **24**, 373-389.

TRUESWELL, J. C., & HAYHOE, M. M. (1993). Surface segmentation mechanisms and motion perception. *Vision Research*, **33**, 313-328.

NOTES

1. An ANOVA without arcsine transformation also showed a significant effect of noise [$F(4,8) = 5.16, p < .05$]. The effect of depth [$F(1,2) = .202$] and its interaction with noise [$F(4,8) = 1.50$] were not significant ($p > .05$).

2. One may argue that some sort of retinotopic integration can occur in the experimental situation in the present study and that the discriminations could have been made on the form that was perceptible from retinotopic integration but not amodal completion. With a 300-msec integration window, about 29% of the face in Experiment 1 and about 57% of the pattern in Experiment 3 could have been defined retinotopically, and these extents might have been sufficient for the discrimination task. Although there is no conclusive evidence to argue against this criticism, it is unlikely that the retinotopic integration played a major role in the present study, because the experimental situation did not favor possible retinotopical integration. Morgan, Findlay, and Watt (1982) have summarized the circumstances that produce retinotopical integration as follows: presence of an independent tracking point, fast stimulus velocity, no bright contour at the slit boundary, and dim background illumination. In contrast, in the present situation, there was no tracking point, the stimulus velocities were slow relative to those in other studies, there was a clear contour at the boundary of the aperture, background illumination was not dim, and the aperture was masked with random dots immediately after presentation in Experiment 1.

3. An ANOVA without arcsine transformation showed a significant interaction between noise and type of response [$F(4,8) = 4.25, p < .05$]. The main effect of depth [$F(1,2) = 12.0$] and its interaction with noise [$F(4,8) = 0.181$] were not significant ($p > .05$).

4. An ANOVA without arcsine transformation did not show a significant effect of depth [temporal, $F(1,2) = 4.71, p > .05$; spatial, $F(3,6) = 0.148, p > .05$].

(Manuscript received January 13, 1997;
revision accepted for publication December 10, 1997.)