PERGAMON

Vision Research 38 (1998) 2081–2085

Vision Research

Stereo capture: local rematching driven by binocularly attended 3-D configuration rather than retinal images

Xin-Nian Wu *, Qing Zhou, Xiang-Lin Qi, Yun-Jiu Wang

Laboratory of Visual Information Processing, Institute of Biophysics, Academia Sinica, Beijing 100101, China

Received 1 August 1995; received in revised form 5 August 1996; accepted 3 December 1997

Abstract

Previous explanations for stereo capture were mainly based on the low-level perceptual processing of binocular stereopsis which usually shows that one pair of retinal images corresponds to only one 3-D perceptual configuration. Stereo capture, however, may encounter multiple perceptual configurations due to the matching ambiguity of wallpaper elements that may not be solved merely by bottom-up processing of the retinal stimuli. The present study suggests that binocular attention plays an important role in stereo capture by way of selecting and enhancing a perceptual configuration that is often ambiguous without attention involved. Stereo capture results from wallpaper's local rematching driven by binocularly attended 3-D configuration rather than retinal images.

Keywords: Stereo capture; Wallpaper effect; Binocular attention

1. Introduction

The intriguing phenomenon known as stereo capture was originally observed in a Kanizsa-type stereogram that was imposed on a wallpaper pattern (Fig. 1) [1,2]. When such a stereogram was binocularly fused, the dots enclosed within the subjective square appeared standing out in front of the background, while the dots outside remained flush with the background.

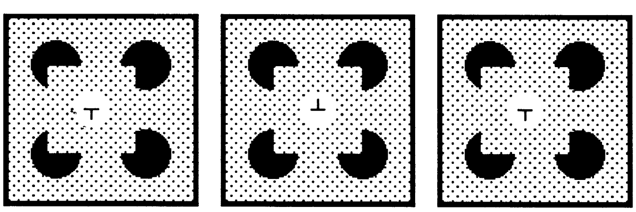

Fig. 1. Stereo capture can be obtained from the two left or two right images in the Kanizsa stereogram by converging or diverging eyes respectively. The appropriate capture in the following stereograms can be seen in the same way. Nonius lines are superimposed on the depth plane of the subjective square to monitor binocular fixation.

* Corresponding author. Fax: +86 10 62027837; e-mail: qixl@sun5.ibp.ac.cn.

0042-6989/98/$19.00
PII: S0042-6989(97)00421-5

The explanations for stereo capture remain controversial. Ramachandran and Cavanagh [1,2] emphasized the role of the subjective stereo surface in stereo capture, which was considered to serve as an important prerequisite. Mather [3] argued by demonstrations that outline figures without subjective contours could also produce stereo capture. Their explanations may involve some form of spreading signal from the subjective square [4]. In contrast to the above explanations, Vallortigara and Bressan [5] proposed that 'stereo capture arises as the solution to a conflict between information provided by retinal disparity and occlusion'. It was shown by Watanabe and Cavanagh [6] that depth capture can also be seen with random dots held in front of or behind an isoluminant patch.

In this paper, the focus of our explanation is concentrated on the additonal role of observers' binocular attention, besides that of the 3-D configuration imposed on wallpaper elements. Binocular attention here is defined as the focus of attention on a binocularly-fused configuration of a considerable extent in 3-D space. This is different from location-based or feature-specific attention in 2-D space [7]. The approach to the role of binocular attention in stereo capture was motivated by the phenomenon noted in our observations that the same Kanizsa stereogram can induce stereo capture as well as wallpaper effects (see also [1,2,5]), implying that this effect may not be induced merely by the retinal images.

Demonstrations: Stereo capture depends on binocularly attended configuration rather than retinal images.

The following demonstrations were designed to show the ambiguity of a configuration imposed on wallpaper and the role of binocular attention in stereo capture. Observers' binocular attention was simply drawn by the experimenter's oral instructions.

2. Method

Five subjects with normal binocular vision took part in the demonstrations. Among them, two were experienced and three were naive in the purpose of the study. All of them can free-fuse stereograms by converging (that is, the right eye's image appears on the left and the left eye's image on the right) or diverging eyes (the right eye's image on the right and the left eye's image on the left).

Stimuli were generated on the standard 14-in color screen of a personal computer. Subjects were instructed to free-fuse stereograms by converging eyes. At the distance of 40 cm, the visual angle subtended by each eye's image was 5 deg arc and the horizontal interval between adjacent wallpaper elements was 11 min arc. The disparity of the contours bounding the capturing surface was 11 min arc. The stereograms presented in the paper consist of three images aligned side-by-side to obtain the appropriate sign of depth for readers by converging or diverging eyes.

In order to dissociate the effects of binocular attention from that of vergence eye movements on stereo capture, a small nonius fixation figure was superimposed on stereograms to monitor eye vergence states [8,9]. During observations, subjects were asked to binocularly fixate on the nonius lines. Under proper binocular vergence, the vertical nonius lines should be in register, otherwise out of register.

2.1. *Effects of binocular attention to local or global configuration on stereo capture*

The most salient stereo feature in the Kanizsa stereogram in Fig. 1 is the subjective square floating in front of the background, which was ever considered as an important prerequisite for stereo capture. However, if subjects were asked to pay attention merely to the front subjective square, ignoring the four 'pacmen' on the background, they reported seeing wallpaper effect instead of stereo capture: the dots within the whole stereogram were pulled up in depth from the background. If they were instructed to binocularly attend to both the subjective square in front and the 'pacmen' on the background, all of them saw stereo capture.

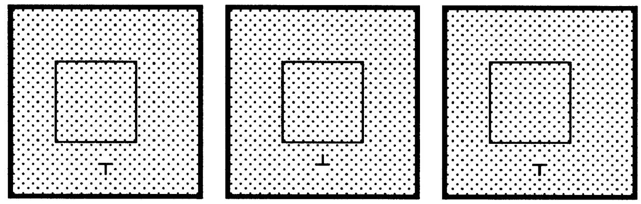

Fig. 2. The outline square in front is ambiguous in stereo capture. The ambiguity can be resolved by observers' binocualr attention.

The results showed that attention to the global 3-D structure of the fused Kanizsa stereogram was in favor of stereo capture, and attention to a local surface of the subjective square facilitated the occurrence of wallpaper effect.

2.2. *Solving the ambiguity of a configuration in stereo capture by binocualr attention*

Some outline configurations may be ambiguous in stereo capture. In Fig. 2, if the subject's attended configuration was a square surface standing out from the background, stereo capture was seen; otherwise if subjects were hinted to pay attention to a square wireframe in front of the background, all of the dots including those within the square were seen to remain flush with the background. The dots both inside and outside the outline square can also be captured to the depth plane of the outline square, if subjects pay attention only to the front outline square.

This demonstration indicated that binocular attention can solve the ambiguity in stereo capture.

2.3. *Selecting a capture structure from a complex configuration by binocualr attention*

Three capture structures could be induced by the stereogram in Fig. 3. First, the dots within the small outline circle were pulled up in depth from the background and those outside the subjective disk on the-

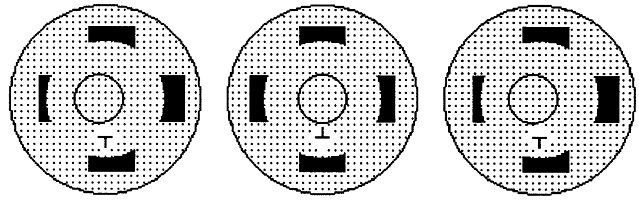

Fig. 3. Stereo capture induced by a more complex 3-D perceptual configuration: a small outline square in front of the background occludes a subjective disk on the background, which in turn occludes a black cross behind the background. The perceived shape of capture depends on the partial configuration which observers binocularly attend to.

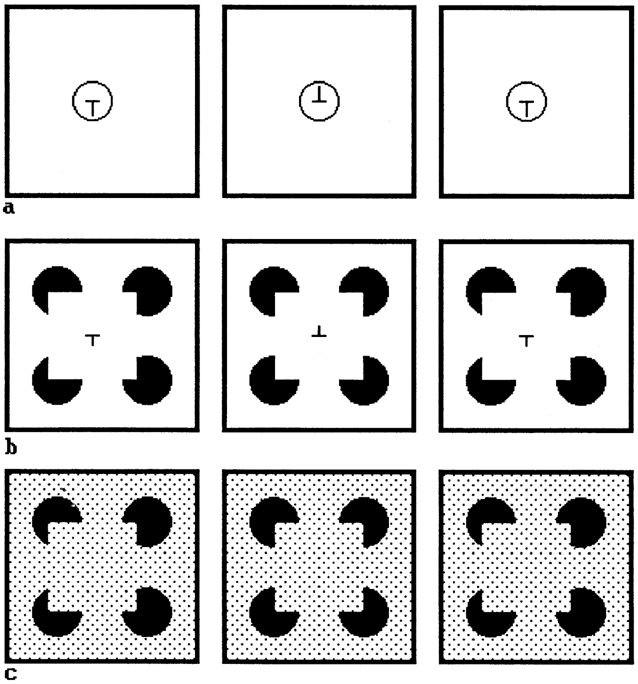

Fig. 4. Stereograms in Fig. 4a, 4b and 4c were used for binocular fixation, cueing attention to different parts of a configuration and testing the role of attention in stereo capture respectively.

background were captured behind. Second, only the dots enclosed in the outline circle were captured forward, and the rest of dots remained flush with the background. Finally, only the dots outside the subjective disk were captured behind, and the remaining dots including those within the small circle lay on the background. The observations showed that the capture structures could be selected by binocular attention.

All of the demonstrations mentioned above may imply that stereo capture is induced by the attended 3-D configuration rather than the retinal images, which remain almost the same when subjects keep their steady binocular fixation on the nonius lines.

2.4. Cue-test experiment: effects of binocular attention cued by a dynamic Kanizsa stereogram

Binocular attention in above demonstrations was simply drawn by oral instructions. The following experiment was designed to examine the effects of attention cued by local or global features in a Kanizsa stereogram which were made flashing.

3. Method and procedure

Three subjects took part in the cue-test experiment. One was experienced (the subject X.W) and the other two were naive. Stereograms were free-fused by converging eyes. Three dynamic modes of a special Kanizsa stereogram without wallpaper elements (Fig. 4b) served as cue stereograms to draw attention. The first and the second modes were flashing modes. The flashing frequency was 4 Hz. In the first flashing mode, the global mode, both the subjective square in front and the four sectored disks on the background were flashing for the purpose of drawing binocular attention to the global 3-D configuration. This could be done by alternately presenting two stereograms at the same place: the first stereogram was the same as that in Fig. 4b, the second was with four completed disks, the diameter of which was about 3 min arc bigger than that of the sectored disks in Fig. 4b. The centers of the sectored disks in the first stereogram coincided with those of the completed disks in the second stereogram. In the second flashing mode, the local mode, only the subjective square in front was flashing. To obtain the local mode, the diameter of the completed disks in the second stereogram remained the same as that in Fig. 4b. The local mode was intended to capture binocular attention to the local subjective square surface. The third mode served as the control mode without any feature flashing. Every subject continuously took a group of 60 trials. Every trial begins with a binocular fixation stereogram (Fig. 4a) which was kept until a key was pressed by subjects after binocular fusion established. Then one of the three modes of the Kanizsa cue stereogram (Fig. 4b) appeared for 3000 ms which was randomly chosen. The subjects were instructed to keep their binocular fixation on the nonius lines and pay attention to the flashing features in the cue stereogram if any. Immediately after the cue stereogram turned off, the test stereogram was presented for 160 ms (Fig. 4c), which was the typical Kanizsa stereogram for stereo capture without any feature flashing. The only task for the subject, after a trial, was to report whether stereo capture or wallpaper effect were seen, by pressing a corresponding key. This was a forced choice process.

4. Results and analysis

The results of the cue-test experiment are shown in Table 1. A 2×2 χ^2-square test was first applied to the data obtained from the response of each subject. Since there was no significant difference among the responses of individual subjects under each of the three modes, the same test could then be applied to the sum of the three individual responses (see Table 2). Compared with the responses under control mode, we can see that the total probability of perceiving either stereo capture or wallpaper effect showed a significant difference between global and local mode ($P < 0.001$). The statistical analysis indicated that the global mode facilitated the occurrence of stereo capture and the local mode was in favor of wallpaper effect.

Table 1
The responses with binocular attention cued by the dynamic Kanizsa stereogram

	Global mode			Local mode			Control mode		
	Freq. (G)	Capture	Wall-paper	F req. (L)	Capture	Wall -paper	F req. (C)	Capture	Wallpaper
X.W.	24	20	4	15	2	13	21	12	9
Q.S.	17	15	2	16	5	11	17	14	3
X.D.	23	15	8	19	1	18	18	11	17
(Sum)	64	50	14	50	8	42	56	37	29

The total trials were 60 (= freq. G + freq. L + freq. C).

The experiment may provide further evidence for our argument that binocular attention cued by different dynamic modes of a stereogram plays an important role in stereo capture, which is induced by binocularly attended 3-D configuration rather than retinal images.

5. Discussion

It seems that stereo capture can occur automatically without voluntary binocular attention. This situation may happen when observers take the first glance over a novel 3-D configuration imposed on wallpaper elements. The global disparate features, such as both the subjective square and the 'pacmen' in the Kanizsa stereogram, may automatically draw binocular attention to the global 3-D configuration and result in stereo capture. It might be the reason why the probability of perceiving stereo capture with the control mode showed no significant difference from that with the global mode (see the data in Table 1). Binocular attention to the global 3-D configuration was involved in fact. The ratio of the length of the subjective contours to the entire length of the subjective square can influence the perceived salience of the subjective square [10], which in turn influences observers' binocular attention. For a prolonged period of observation, however, observers' attentional states may fade away or involuntarily shift to a salient local feature, such as the subjective square. The change of the attentional states can cause stereo capture broken. We believe that binocular attention is required in general for stereo capture.

It was shown that binocular vergence can influence the assignment of stereo correspondence [11] and may play a role in stereo capture. The technique of the nonius lines was used to estimate observers' vergence states by monitoring the amplitude and the direction of the offset between the upper and the lower segmant of the nonius lines. It can enable subjects to minimize the amplitude of vergence eye movements during steady fixation. For the brief presentation (160 ms) of the test stereogram, any eye movements including vergence should be forbidden. Only covert shifts of binocular attention were allowed. Small eye movements, however, may be unavoidable during the prolonged presentation of the cue stereogram even with steady binocular fixation. We noted that the amplitude of the offsets of the nonius lines in this case was much smaller than when subjects voluntarily shifted their vergence from the fixation plane of the subjective square to the 'pacmen' on the background. The amplitude can be inferred from the measurements of the thresholds for dichoptic vernier acuity [9]. In natural viewing condition, eye movement in 2-D space is one form of visual attention. It is plausible that vergence eye movements are also bound to follow observers' binocular attention to a 3-D configuration of a considerable extent in depth. It was suggested that vergence eye movement was unlikely responsible for stereo capture. Two planes in different depths were simultaneously observed in stereo capture, an effect that would be impossible to achieve by simply using vergence alone [1,2].

Table 2
2×2 χ^2-square test applied to the data in Table 1

	Global versus local	Local versus control	Global versus control
X.W.	18.4 ($P<0.001$)	6.63 ($P<0.01$)	2.725 ($P>0.1$)
Q.S.	11.19 ($P<0.001$)	8.81 ($P<0.01$)	0.23 ($P>0.6$)
X.D.	17.91 ($P<0.001$)	13.3 ($P<0.001$)	0.086 ($P>0.5$)
Total	43.49 ($P<0.001$)	27.13 ($P<0.001$)	2.19 ($P>0.25$)

There was an analogy between stereo capture and motion capture [2,12], which can also be mediated by attention [13,14]. However, there is a significant difference between the two effects. Stereo capture is the outcome of local rematching of wallpaper elements on the basis of local surfaces of the imposed configuration. The process takes place for the entire wallpaper, element-by-element, including those on the background. The problem of the interocularly unpaired elements on the background can be solved by using the occlusion constraints [15,16]. For motion capture, however, the signal from a salient feature can simply be attributed to the enclosed elements, without need of element-by-element matching. We use the term of 'local rematching' to emphasize the essential difference of stereo capture from motion capture as well as wallpaper effect. More-

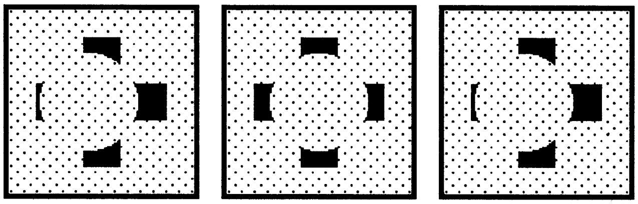

Fig. 5. Observers can see stereo capture under either uncrossed or crossed disparities. The converger can see, by free-fusing the two left images, that a white subjective disk with the dots inside standing out in front and occluding a black cross on the background. By free-fusing the two right images, however, the dots within the disk are captured behind the background (seen through a transparent cross).

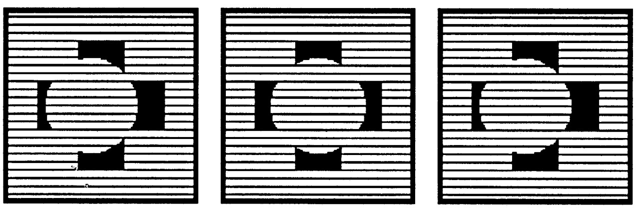

Fig. 6. The four 'pacmen' imposed on the wallpaper are shaped to form a tilted subjective square. When fused, the part of horizontal lines within the square will be captured to the tilted subjective surface and the remaining parts, outside the square, stay at the background.

over, the argument concerning stereo capture with uncrossed disparities can also be resolved based on the local rematching requirement. Ramachandran [2] ever reported an unexpected 'illusion' in stereo capture. When uncrossed disparities were introduced into the cut sectors in the Kanizsa stereogram, the 'porthole' illusion was seen instead of perceiving the expected illusory square falling behind the background. We realized by careful analysis that the 'porthole' illusion may be the only reasonable perceptual structure satisfying the local rematching requirement. It is well known that binocular matches on frontoparallel planes require the corresponding features to be identical in shape. If the whole subjective square did fall behind the background, the corresponding 'pacmen' in the two eye images could hardly match each other on the background because they are rather different in shape (see in Fig. 1). We can show that any desired shape of stereo capture can be achieved provided that the corresponding perceptual configuration can be constructd and the wallpaper texture is available for the configuration. The disparity can be crossed or uncrossed (Fig. 5). The capturing surface can be frontoparallel (Fig. 1), tilted (Fig. 6) or even curved ones (Fig. 7).

Binocular stereopsis depends critically on the solution to the matching problem arising from the so-called global stereopsis which was considered as a 'cognitive impenetrable module' in early visual processing without top-down influences [17]. The present study implies that binocular attention may also be involved in binocular matching process especially when the correspondence is ambiguous.

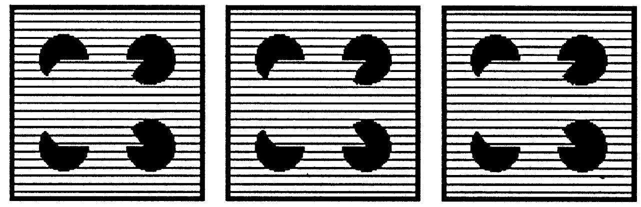

Fig. 7. A curved subjective square and the corresponding capture can be perceived.

Acknowledgements

The work is supported in part by NSFC grant No. 39710106. We would like to thank Manfred Fahle and two anonymous referees for helpful suggestions and comments on previous versions of the manuscript.

References

[1] Ramachandran VS, Cavanagh P. Subjective contours capture stereopsis. Nature Lond 1985;317:527–8.

[2] Ramachandran VS. Capture of stereopsis and apparent motion by illusory contours. Percept Psychophys 1986;39:361–73.

[3] Mather C. The role of subjective contours in capture of stereopsis. Vis Res 1989;29:143–6.

[4] Mitchison GJ, Mckee SP. Interpolation in stereoscopic matching. Nature 1985;315:402–4.

[5] Vallortigara G, Bressan P. Occlusion, transparency and stereopsis: a new explanation for stereo capture. Vis Res 1994;34(21):2891–6.

[6] Watanabe T, Cavanagh P. Depth capture and transparency of regions bounded by illusory and chromatic contours. Vis Res 1992;32:527–32.

[7] Rossi AF, Paradiso MA. Feature-specific effects of selective visual attention. Vis Res 1995;35(5):621–34.

[8] Krol JD, van de Grind WA. Binocular depth mixture: an artifact of eye vergence? Vis Res 1986;26:1289–98.

[9] Fahle M. Psychophysical measurement of eye drifts and tremor by dichoptic or monocular vernier acuity. Vis Res 1991;31:209–22.

[10] Lesher GW, Mingolla E. The role of edges and line-ends in illusory contour formation. Vis Res 1993;33:2253–70.

[11] Mallot HA, Bideav H. Binocular vergence influences the assignment of stereo correspondences. Vis Res 1990;30:1521–3.

[12] Ramachandran VS, Cavanagh P. Motion capture anisotropy. Vis Res 1987;271:97–106.

[13] Culham JC, Cavanagh P. Motion capture of luminance stimuli by equiluminous color gratings and by attentive tracking. Vis Res 1994;34:2701–6.

[14] Ramachandran VS. Motion capture and Visual attention: comments on Culham and Cavanagh (1994). Vis Res 1996;36:77–8.

[15] Shimojo S, Nakayama K. Real world occlusion constraints and binocular rivalry. Vis Res 1990;30:69–80.

[16] Nakayama K, Shimojo S. Da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. Vis Res 1990;30:1811–25.

[17] Julesz B. Early vision is bottom-up, except for focal attention. Cold Spring Harbor on Quantitative Biology, Volume LV, 1990.