Two processes in stereoscopic apparent motion

Hiroyuki Ito *

Department of Visual Communication Design, Faculty of Design, Kyushu Institute of Design, 4-9-1 Shiobaru, Minami-ku, Fukuoka-shi 815-8540, Japan

Received 2 October 1996; received in revised form 10 March 1998

Abstract

This study investigated the human ability to discriminate the motion direction of sequentially presented depth patterns produced by random-dot stereograms. The stereoscopic (cyclopean) patterns used here consisted of 256 rectangle patches, each of which had an alternative depth position (near or far). Two successive frames of correlated depth patterns made impressions of lateral motion when the pattern position in the second frame shifted laterally. The density of the patches that were near was varied. The $D_{\text{max}}$ that was measured using the 2AFC method was short when the density was high. The effect of depth reversing in the second frame was also tested. Under low density conditions, the performance was still good against reversing 3-D polarity. However, when the density was high, with depth reversal, motion in the reversed direction was perceived. Reversed motion was observed more often when SOA was small and when the density of near patches was near $1/2$. Two strategies seem to exist in stereoscopic motion detecting: a polarity-independent process which matches figures, ignoring their depth polarity, and a polarity-dependent process which operates locally, ignoring 2-D shapes. The latter suggests the existence of a passive process in stereoscopic motion. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Stereoscopic (cyclopean) motion; Random-dot stereogram; Higher-order motion; Reversed-phi; $D_{\text{max}}$

1. Introduction

The human visual system can detect motion, using edge or form information defined by differences in luminance, color, texture or binocular disparity (Cavanagh, Arguin & von Grunau, 1989). Luminance- or color-defined motion is often called first-order motion, whereas motion defined by other attributes is called second-order motion (Cavanagh & Mather, 1989; Cavanagh, 1991). Apparent motion produced by changes in the binocular disparity distribution has been thought of as one type of second-order motion, and is called stereoscopic (cyclopean) motion (SM). Julesz and Payne (1968) first developed a technique of presenting stereoscopic apparent motion, i.e. the successive presentation of random-dot stereograms (RDS) which represented 3-D forms correlated between frames, although the dot distributions were uncorrelated. They showed that horizontal and rotational motion of stereo-defined forms can be perceived without correlation in the luminance domain. The present study investigated the human ability to discriminate the motion direction of sequentially presented depth patterns produced by RDS.

According to Cavanagh (1991, 1995), motion processing has two different aspects. One is the active process, attentional feature tracking, and the other is the passive process, non-cognitive motion sensing. Within the first-order (luminance) domain, ‘long-range motion’ and ‘short-range motion’ processes seem to correspond to the two qualitatively different processes noted above, respectively, although the short-range and long-range distinction is questioned by Cavanagh and Mather (1989) and Cavanagh (1991). Active motion processing deals with visible features, e.g. shapes or edges, but it sometimes seems insensitive to their polarity. For example, even when luminance contrast is reversed between frames, apparent motion, called epsilon motion, occurs if the shapes are clearly visible and if the displacement between frames is more than 15 min arc (Kolers & von Grunau, 1976). Such a process can gather information over a large extent of space and time. The spatial limit of displacements for long-range
motion perception exceeds several degrees, and the temporal limit of inter-stimulus intervals (ISI) for motion perception seems to be more than 80 ms. On the other hand, passive motion processing, called motion sensing, functions even when there is no visible contour. For example, the visual system can detect motion from successively presented random patterns (Anstis, 1970; Braddick, 1974). This process seems to be polarity dependent because, when the contrast is reversed in the second frame, motion in the reversed direction is seen (‘reversed phi’ movement, Anstis, 1970). Motion sensing in the luminance domain seems limited in narrow spatial and temporal extents. In the case of random-dot stimuli, if the displacement is more than 15 arc min, or if the ISI exceeds 80 ms, motion is not observed.

The existence of the two qualitatively different strategies noted above in second-order motion was suggested by Smith (1994). He investigated one kind of second-order motion (contrast-modulated motion) and separated ‘feature-based mechanisms’ from ‘intensity-based mechanisms’ by introducing ISI and feature masking. This indicates that there are three motion-detection mechanisms, i.e. two which are intensity-based (including first-order motion) and one which is feature-based. Recently, Lu and Sperling (1995) also suggested the existence of ‘third-order’ mechanisms in the visual system, distinct from other second-order motion processes. What they called ‘third-order’ motion is thought to be detected by a feature-tracking strategy, not energy computation. This conclusion is quite similar to that of Smith (1994).

As for SM, contradicting results have been reported. Cavanagh (1995) suggested that there is no low-level processing for SM, i.e. detecting SM is an active process, showing that SM was barely seen without attentional tracking. Lu and Sperling (1995) also concluded that SM is detected by a ‘third-order’ mechanism. However, there are a few reports which show that SM is not an active process. These show that SM adapts and motion aftereffects occur, although this requires a rather long viewing period (Patterson, Bowed, Pinney, Pohnndorf, Barton-Howard & Angilletta, 1994; Bowed, Rose, Pinney & Patterson, 1996; Patterson & Becker, 1996). They assert that their phenomenon yields evidence indicating that SM is a sensory process, not a cognitive one like attentional feature tracking, which does not adapt.

One of the purposes here is to examine the importance of figural structure in stereoscopic apparent motion. Easily recognizable figures may provide a basis for attentional tracking. To test this, arrays of randomly positioned rectangle patches defined by disparity were used, i.e. random-depth patterns produced by assigning small patches ‘near’ or ‘far’ values, whereas ordinary random-dot patterns are produced by assigning the dots ‘black’ or ‘white’ values. This stimulus composi-

tion is similar to that of Patterson, Donnelly, Pinhe

n, Nawrot, Whiting and Eyle (1997). Two frames of ran
dom-depth patterns created stereoscopic apparent mo

tion. Using random patterns has the advantage of quan
titative control of the figural structures. When there are only a few near (or far) patches, it will be easy to recognize the ‘figure’ against the ‘ground’ area and to match the figural patterns between frames. However, when half of the patches are near, it is difficult to recognize the figure and to match the patterns between frames without false correspondence. This is similar to the procedure used by Ramachandran and Anstis (1983) and Sato (1990) who showed that an escalation of $D_{\text{max}}$ occurs according to the decrease of random-dot density. Boulton and Baker (1993a) also showed that the ‘quasi-linear’ and ‘nonlinear’ mechanisms are distinct with the density of the stimulus elements and that the ‘nonlinear’ process is favored by sparsely populated stimuli. They successfully controlled the qualitative change of motion processes using only one parameter, i.e. density. At present, stimulus patterns used in SM experiments have been quite simple, such as discs (Patterson, Hart & Nowak, 1991; Pinney, Wilson, Hays, Peters & Patterson, 1994) or gratings (Julesz & Payne, 1968; Chang, 1990; Patterson et al., 1994; Ito, 1997), except for Patterson et al. (1997). This may be the one reason why the $D_{\text{max}}$ of the SM seems larger than that of the first-order motion and why the SM seems to be an active process.

The other factor tested here was the depth reversal in the second frame. Even if the depth polarity is reversed, i.e. near patches become far ones in the second frame and vice versa, one can recognize the original figural structure in the reversed frame when the pattern is simple enough, resulting in epsilon motion in the depth domain. Apparent motion between near and far planes defined by stereopsis has been reported on by Cavanagh et al. (1989), who used a stereo-defined disc, i.e. the simplest pattern. If SM detection functions as a pattern matching or feature tracking process, easily recognizable figures will cause good SM perception even when the depth is reversed. However, when the figural structures in the stimuli are not easily recognized in the depth-reversed patterns, the motion detection performance will decrease and cause random motion perception or none at all.

2. Experiment 1

2.1. Method

2.1.1. Subject

Three subjects participated in the experiment. One was the author and the others were students of Kyushu Institute of Design who did not know the purpose of
the experiment. All had normal or corrected-to-normal visual acuity and good stereo vision.

2.1.2. Apparatus and stimuli

Two frames of RDS were generated by a microcomputer (SHARP CZ-644C) and presented on two CRT monitors (SHARP CZ-614D) refreshing at 55 Hz. The vertical and horizontal-sync signals from the computer were distributed to the displays in parallel, synchronizing their refresh. Each display presented a corresponding pattern to each eye with green dots (4.7 cd/m²) with a dark background (under 0.01 cd/m²). Subjects stereoscopically viewed them through mirrors at a distance of 65 cm (see Ito, 1997).

Fig. 1 shows a schematic illustration of the cyclopean stimuli. The size of the screen was 26.5 cm (horizontally) x 18.5 cm (vertically) and the resolution of the screen was 512 x 512 pixels. Horizontal pixel length (2.7 arc min) was not the same as the vertical length (1.8 arc min). The screen was treated as 16 x 16 rectangle patches. Each patch had a 32 x 32 pixel area and included 16 randomly distributed dots, each of which consisted of 2 x 2 pixels. One of the two disparity values (0 or 5.4 arc min in the crossed direction) was randomly assigned to each patch. All 16 dots within one patch had the same disparity value. When these parameters were not appropriate, one could not see a valid depth pattern. For example, when the size of a patch was not large enough, compared to the phenomenal depth separation or when the dot density was too low, grouping of dots across patches occurred. As a result, two transparent planes separated by depth were perceived while the depth pattern was not seen.

The depth pattern of the first frame was presented with a random offset along the horizontal axis, i.e. the depth edges in the patterns was not always presented in the same place on the screen. However, irrespective of the offset, bright dots and a depth pattern filled the whole display area. After the first frame presentation, the depth pattern was shifted to the right or the left in the second frame. According to the shift, some patches went out or appeared at the lateral edges of the screen. Luminance dots were static during each frame presentation. The duration of each frame was 182 ms (ten times monitor refresh) and ISI was zero. Although the depth patterns were correlated between the two frames, dot patterns were not correlated. When one looked at the successive frames with one eye, random motion of random dots filling the screen was seen, without any visible figures.

The density of near patches (having crossed disparity) varied (1/16 or 1/2). In order to measure $D_{\text{max}}$, the horizontal displacement between the frames was also varied, i.e. 16, 32, 48, 64, 80, 96 or 112 pixels (from 43 to 302 arc min). The third variable was depth polarity. Strictly speaking, ‘polarity’ as it is used here does not represent a disparity value itself, i.e. crossed or uncrossed, but a relative depth position of a patch (convex or concave), against its surroundings. In half of the trials, the pattern shifted left or right in the second frame without changes of depth polarity. In the other half, the depth of the second frame was reversed, i.e. patches having crossed disparity in the first frame had zero disparity in the second frame, and vice versa. Fig. 2 shows the difference of the two conditions. When the
density was 1/2 with a depth reversal, it was difficult to identify the first pattern in the second frame.

2.1.3. Procedure

A small ‘+’ was located in the center of the display as a fixation point, i.e. having zero disparity. After a beep sound, two frames of RDS were successively presented. The subject’s task was to choose ‘left’ or ‘right’ according to their perceived direction of SM.

The experiment was separately conducted for each density condition and for each subject. Under each density condition, five experimental sessions were conducted after one training session. One session included ten repetition trials for each combination of seven displacement \( \times \) two polarity (normal and reversed) conditions. Two directions of pattern shifts were randomly assigned with the same likelihood. When subjects could not discriminate motion directions, the percentages of correct responses was around 50. The order was randomized within each session. No feedback was given.

2.2. Results and discussion

The results obtained under density conditions of 1/16 are shown in Fig. 3. Under normal conditions, all subjects discriminated the motion direction almost perfectly when displacement was small. \( D_{\text{max}} \) (75% correct) was 3.5, 3.7 or 2.3° for HI, TS or IO, respectively. Even under depth-reversal conditions, the performance was still good. It is clear then that the ‘figure-ground’ relationship was preserved and subjects could match the figural patches between frames even when the depth polarity was reversed. This phenomenon seems to correspond to epsilon motion in the luminance domain. 2-D feature tracking seems to be an effective strategy in motion direction discrimination of stereo-defined forms, although the effect of cognition might be involved. Under density conditions of 1/2 (see Fig. 4), when the depth polarity was preserved, the percentages of correct responses decreased with an increase in displacement. As shown in Fig. 5, the \( D_{\text{max}} \) seems much smaller than that under a density condition of 1/16 (1.9, 2.5 or 1.1° for HI, TS or IO, respectively). The density of the figural patches then may have limited the displacement of possible motion detection. The \( D_{\text{max}} \) of SM reported by Phinney et al. (1994) is 5° under 116 ms duration and zero ISI condition, which is rather large compared to the results here. Apart from the fact that their procedure used for measuring \( D_{\text{max}} \) was different to the procedure used here, another reason for the difference may be the figural simplicity of their stimulus (i.e. a disc). The ease of recognizing the figural structure may be one factor in defining \( D_{\text{max}} \). Another possible explanation for the difference of \( D_{\text{max}} \) between density conditions of 1/2 and 1/16 is that edge information after spatial pre-filtering (e.g. zero-crossings in the depth domain) is matched between frames, and that sparse patches contribute to large jumps without false correspondence. Morgan (1992) proposed such a strategy for the apparent motion of luminance-defined random dots. Eagle and Rogers (1996), Eagle and Rogers (1997) and Morgan, Perry and Fahle (1997) also support the idea. This strategy can predict the effect of element density on \( D_{\text{max}} \) under normal conditions, but can not explain the results under density conditions of 1/16 with depth reversal where edges of opposite signs correspond. Applying Morgan’s theory to depth domain also predicts the effect of patch size on \( D_{\text{max}} \) under density conditions of 1/2. This prediction is now being tested in my laboratory.

The performance under density conditions of 1/2 with depth reversal was surprising. The percentages of

Fig. 3. Percentages of correct responses under 1/16 density conditions. Open or filled circles indicate the performance under the normal or reversed conditions, respectively. Even under depth-reversed conditions, the performance is only a little below that under the normal conditions. The error bars in the figure (and in following figures) indicate 95% confidence intervals.
The results suggest that, besides a polarity-independent pattern matching (or feature tracking) system, there is a polarity-dependent motion sensing system in SM processing. The performance curves of the depth reversed conditions are like the reversals of those under the normal conditions when the density is 1/2. This may inversely indicate that the polarity-dependent motion sensing process also greatly contributes to motion detection under normal conditions with the density of 1/2. On the other hand, under density conditions of 1/16, the performance for almost all displacements in the reversed condition is a little worse than the normal condition. The poorer performance in the reversed condition could be due to the fact that the low-level (polarity-dependent) and high-level (polarity-independent) processes are in conflict.

3. Experiment 2

The results of Experiment 1 show that when the depth pattern is simple enough to extract the ‘figure’ in each frame, pattern matching ignoring depth polarity can be used as a strategy to determine the motion direction. When the pattern is too complex to find itself in the depth reversed frame, a polarity-dependent motion sensing process works effectively. Experiment 2 investigated the effect of density on perceived motion direction under depth reversed conditions. Does the reversed motion phenomena occur only under density conditions of 1/2?

3.1. Method

Two subjects from Experiment 1 participated in Experiment 2. Normal and reversed depth conditions were also used here. The number of near patches in the first frame was varied, i.e. 8, 32, 56, 80, 104 or 128 out of 256 patches (these correspond to densities from 1/32 to 1/2). The displacement between frames was fixed to 32 pixels (86.4 arc min), at which percentages of correct
responses were highest under normal density conditions of 1/16 and lowest under reversed density conditions of 1/2 in Experiment 1. Two ISI conditions were also used, 0 and 182 ms. During ISI, only a fixation cross was presented. It was shown that inserting ISI decreases the interference of luminous dot motion while detecting SM (Ito, 1997). One experimental session included 120 trials, six density × two polarity (normal and reversed) × ten repetitions. Under each ISI condition, five sessions were conducted. The order was randomized within each session. The other method was the same as in Experiment 1.

3.2. Results and discussion

Fig. 6 shows the results of Experiment 2. Under normal conditions, both subjects discriminated motion directions almost perfectly, irrespective of increasing density, i.e. the number of near patches. Under depth reversed conditions, however, perceived motion direction changed according to the increasing number of near patches in the first frame. This may indicate that the dominant strategy of motion detection shifted from pattern matching to motion sensing as the figural patches became denser. The element density seems to be a good parameter for controlling the relative strength of the two processes, as was shown by Boulton and Baker (1993a). For both subjects, under denser conditions, when ISI was 182 ms, reversed motion was weak compared to the results obtained under ISI conditions of zero. For subject TS, under lower density conditions, percentages of correct responses were higher when ISI was 182 ms. Inserting ISI decreased the percentages of reversed motion, that is, the relative strength of pattern matching was increased by ISI. Reversed motion seems to be favored by zero ISI. It is difficult to explain this effect from the point of luminance motion interference because even if the interference exists when ISI is zero, it will never emphasize the reversed motion of stereoscopic depth patterns.

4. Experiment 3

This experiment was conducted to investigate whether the relative strength of the two processes is defined by the ‘number’ or ‘area proportion’ of figural patches. Changing the number also resulted in changing the area proportion in Experiment 2. In Experiment 3, the number of near patches in the first frame was fixed even when the area proportion changed.

4.1. Method

The screen was divided into 4 × 4 rectangle areas. A near patch was placed at a fixed position within each rectangle area with a probability of 1/2. As a result, the number of near patches was fixed at eight in the first frame under all conditions. The area proportion of near patches in the first frame was controlled by changing the size of near patches. As shown in Fig. 7, there are four patch size conditions, 32 × 32, 64 × 64, 92 × 92 and 128 × 128 pixels (1, 4, 9 and 16 times the size of the original patches in area, and these corresponded to area proportions of 1/32, 4/32, 9/32 and 16/32, respectively). The other procedures were the same as in Experiment 2.

4.2. Results and discussion

Fig. 8 shows the percentages of correct responses as a function of the area proportion of near patches in the first frame. Under normal conditions, no effect of the area proportion was found. On the other hand, it is evident that even when the number of near patches was only eight, reversed motion perception occurred under proportion conditions of 1/2 with a depth reversal. Generally speaking, patterns having smaller area and a
convex shape are easily regarded as ‘figures’. When the area proportion is near 1/2, the ‘figure’ in the first frame becomes difficult to find in the second frame, because the original figural patterns change into concave patterns in the second frame and because the concave area in the second frame is not small enough to be regarded as a figure. This may result in increasing the relative strength of polarity-dependent motion processing. It is also possible to argue that the element size has an effect on reversed motion perception. In Experiment 3, the displacement was fixed, i.e. appropriate for reversing the motion of the smallest patches. The relative strength of reversed motion under area proportion conditions of 1/2 cannot be explained by the largest patch size itself.

It is also evident, from a comparison with the results of Experiment 2, that even when the area proportion is the same, a greater number of figural patches strengthens the reversed motion. Increasing the number of near patches leads to the development of complex patterns even when the area is the same. If the pattern is too complex to be recognized, the polarity-dependent strategy will be relatively stronger than the other process.

![Fig. 7. Figural configuration in Experiment 3. The number of near patches was fixed at eight. A near patch was put in fixed positions within 4 × 4 rectangle areas with 1/2 probability. Changing patch size changed the proportion in area without changing the number.](image)

5. Experiment 4

As shown in Experiment 2, reversed motion seems to be favored by short ISI, while the pattern matching process does not. Experiment 4 tested if the manipulation of temporal parameters can isolate the two processes.

5.1. Method

The near patch density was 1/2. The displacement was the same as in Experiments 2 and 3. Frame duration was varied in two, 91 or 455 ms (5 and 25 times monitor refresh, respectively). ISIs were varied in three, 0, 182 or 364 ms (0, 10 and 20 times monitor refresh, respectively). For each combination of two duration × three ISI × two depth (normal and reversed) conditions, there were ten repetition trials within a session. In total ten sessions were conducted after a training session. The order was randomized within each session.

![Fig. 8. Percentages of correct responses as a function of the area proportion of near patches in the first frame. Circles represent the results of Experiment 3. The results of Experiment 2 (under 0 ms ISI conditions) are replotted with squares. Open or filled symbols indicate the performance under the normal or reversed conditions, respectively.](image)
5.2. Results and discussion

Fig. 9 shows the percentages of correct responses as a function of stimulus-onset asynchrony (SOA). Each subject's score under the 91 ms duration with 364 ms ISI condition shows the same performance as that under the 455 ms duration with 0 ms ISI condition, i.e. having the same SOA. When the depth condition was normal, the performance was almost perfect under all the combinations of ISI and duration conditions. When the depth was reversed, however, the percentages of correct responses were under 50; the reversed motion occurred. When the duration was 91 ms with zero ISI (the smallest SOA), quite high percentages of reversed motion occurred. When duration was 455 ms with 364 ms ISI (the largest SOA), however, reversed motion seems weak. The rate of reversed motion under depth reversed condition reflects the relative strength of polarity-dependent motion signals and it seems to decrease (approach the chance level) according to the increase of SOA. As the displacement was fixed in this experiment, larger SOA involves lower velocity. Patterson, Ricker, Megary and Rose (1992) showed that SM detection is governed by velocity. The results obtained here seem to be in agreement with their conclusion. Under larger SOA conditions, however, the pattern matching process seems to operate effectively because the performance under normal conditions does not decrease, whereas reversed motion becomes weak.

6. General discussion

It can be concluded that there are two complementary processes in stereoscopic apparent motion. One is a 2-D figure matching process which effectively functions when the ‘figure’ is clear and when the velocity is low. This process may extract information from disparity-defined edges or shapes and match the 2-D patterns between frames, ignoring depth polarity. Attentional feature tracking may play an important role here. The existence of a polarity-independent process makes it possible to match edges or shapes defined by different attributes. After pattern extraction, the matching process might be common for all attributes (Cavanagh et al., 1989). Only this process can mediate inter-attribute motion because the polarity of an attribute has its meaning only within the domain.

As for the other process, there are two possible explanations; firstly, there is a process of Fourier motion detection in the depth domain. Secondly, each region or edge simply moves to a nearer region having the same disparity value or an edge having the same sign. The difference between them corresponds to that between ‘phase-based and information-based’ motion detection in the luminance domain (Eagle & Rogers, 1996). In any case, the polarity-dependent process operates locally and effectively functions when the velocity is high and when the pattern is complex or dense. This local parallel process may not need pattern global rigidity and some contradiction of motion directions between regions may occur. Actually, such perception was reported by subjects in the cases of depth-reversed stimuli, e.g. some parts seemed to move in the direction opposite to the impression of global motion direction. This process does not seem to be mediated by attention. A passive motion process may exist in SM. The findings here can not be explained by recent models of motion processing, which suggest that detecting stereo-defined motion depends on feature tracking (Cavanagh, 1995; Lu & Sperling, 1995).

The differences in performance between the two above processes, i.e. polarity-dependent and polarity-independent, seem quite similar to those of classical long-range and short-range motion processes, or those of nonlinear and quasi-linear mechanisms in the luminance domain. There are other researches which also suggests such a similarity. The Ternus display holds qualitatively different percepts according to changing
ISI, and the effect was the same for cyclopean Ternus display (Patterson et al., 1991). Boulton and Baker (1993a) showed that perception of a densely packed stimulus was governed by a quasi-linear mechanism and detection of a sparsely populated stimulus was mediated by a highly nonlinear mechanism. The density effect is similar to the results of Experiments 1 and 2. It was also shown that, even using the same stimuli, temporal parameters draw a line between two processes, larger ISI or SOA favored nonlinear processes (Georgeson & Harris, 1990; Boulton & Baker, 1993b). This may correspond to the results of Experiment 4. It seems that SM processing is based on principles similar to those of luminance domain motion sensing and motion perception, although the spatio-temporal scales of SM are several times coarser than those of luminance motion.

Some factors involved in SM remain untested. As shown in Experiment 1, the $D_{\text{max}}$ of SM can be objectively measured. The effects of ISI, duration (or SOA) and element density on the $D_{\text{max}}$ should be investigated. Such data will be used to compare the characteristics between stereoscopic and luminance motion systems. Whether or not a polarity-dependent process exists in other second-order attribute domains, is also worth testing. The technique used here may be useful for such experiments.

Finally, the present paper refers only to relative depth positions within a global region. The depth position can be defined in other ways, absolute disparity values, signs of disparity and perceived positions in a phenomenal 3-D space. Disparity values may change according to vergence change, even if perceived depth positions of objects in a 3-D space do not change. On the other hand, in an experimental situation, it may be possible to keep objects’ absolute disparity values, even when a vergence movement occurs (this may lead to changes in objects’ perceived depth positions). Whether SM is defined by disparity values or perceived depth positions may be an important problem.

Acknowledgements

This research was supported by a grant from Ministry of Education, Science, Sports and Culture (08710050).

References


