Vision Res. Vol. 31, No. 11, pp. 2025–2028, 1991
Printed in Great Britain.
0042-6989/91 $3.00 + 0.00

RESEARCH NOTE

SHORT- AND LONG-RANGE PROCESSES IN STRUCTURE-FROM-MOTION

MIRI DICK, SHIMON ULLMAN and DOV SAGI*
Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 76100, Israel

(Received 21 February 1990; in revised form 7 January 1991)

Abstract—Human ability to detect 3-D structure in an array of 2-D moving dots was tested. Under limited exposure time, we found high detection rates only when the 2-D motion was restricted to the spatio-temporal region of short-range motion. Long-range moving dots failed to produce a strong impression of 3-D structure and yielded only weak detection rates. This result is consistent with the view that the processing of long-range motion is more serial than that of short-range motion.

Structure-from-motion Short-range motion Preattentive detection

INTRODUCTION

Structure-from-motion (SFM, also called "kinetic depth", Wallach & O'Connell, 1953) refers to the ability of the visual system to recover 3-D structure based on 2-D motion information alone. In a typical SFM experiments with discrete points (Green, 1961; Braunstein, 1976; Ullman, 1979; Petersik, 1980), an image of a transparent hollow object, such as a 3-D cylinder, is created by successively presenting its 2-D projections while it rotates, e.g. about its vertical axis. Any single view of the cylinder fails to convey a clear 3-D impression (one such projection is shown in Fig. 1). In the changing image, the sparse dots appear to be glued to the cylinder's surface, and a vivid impression of a rotating cylinder is obtained.

The global percept of the rotating cylinder that is obtained by the displacement of the dots results from their apparent motion (AM). It was suggested (Braddick, 1974; Anstis, 1980; Petersik, 1989) that AM can be described by two mechanisms: short-range (SR) and long-range (LR) motion. It was shown recently (Dick, Ullman & Sagi, 1987) that in visual search tasks SR motion detection can be processed in parallel (preattentively) across the visual field, while the detection of LR motion requires a serial search (is attentive). Thus, if the process that underlies the perception of SFM suffers from the same limitations as parallel search, we might expect perceptual difficulties and evidence for serial processing under pure LR conditions. Alternatively, it is possible that LR motion information is directly available for SFM computations, regardless of attentive limitations.

Earlier studies demonstrated a failure in perceived depth and rigidity of 3-D structure from motion when the motion was of the LR type. In these experiments the observers were presented with an array of random dots simulating transparent rotating objects. The observers' task was to report "depth" of a rotating sphere (Mather, 1989) or "rigidity" of a rotating cylinder (Todd, Akerstrom, Reichel & Hayes, 1988). Both tasks may require higher level decisions involved in recognition of the 3-D object and thus the LR failure may be due to an inability to access these high level operations. We thought to overcome this problem by using a detection paradigm.

In our experiments, we asked whether SFM can be detected when the rotating cylinder is embedded in a background of randomly moving dots and whether this detection can be achieved efficiently when SFM is defined by SR or LR motion only. Although we discuss SR and LR motion in the context of the dichotomy proposed by Braddick (1974), our stimuli for the two types of motion differed only in the spatial displacement of the individual dots between two temporal frames. In all experiments presented

*To whom correspondence should be addressed.

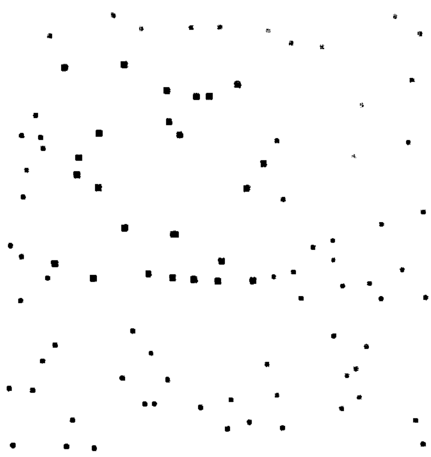

Fig. 1. One frame of the SFM experiment. A single projection of a transparent, hollow, 3-D cylinder against background noise. The larger dots are used here to distinguish the cylinder from the background; in the experiment all dots were identical.

below we made an effort to avoid detection failures due to correspondence problems.

EXPERIMENTAL PROCEDURE

Experiments were performed on the graphic screen of the Symbolic LISP machine with P4 phosphor. The observers watched the bluish screen from a distance of 40 cm. The stimulus we used was a square field of random dots (RD) occupying an area of 10 × 10 deg, out of which a 6 × 5 deg area was designated as the target (the cylinder) area. The cylinder appeared in a randomly-chosen location within the square field. The total number of dots varied from 50 to 200 dots in different experiments. The dots were black, each occupied about 5 min arc of the viewing angle. Four frames were flashed serially on the screen for a duration of 32 msec for each frame. The 3-D cylinder was rotated by a fixed angle between successive frames. The orthographic projection of the cylinder's rotation resulted in horizontal 2-D motions to the left and to the right (the cylinder was transparent, with both the front and back visible). The background was also covered with dots sharing the same density as the average of the projected dots of the cylinder. The dots of the background also moved in a horizontal motion, to the left or the right, in a randomly-chosen velocity around the average speed of the dots on the moving cylinder. The observers' task was to decide whether or not there was a rotating cylinder in the presented random dot array. In each experiment there were 60 trials, in half of the trials the cylinder was present (which required a "yes" response); otherwise there was only the noisy background ("no" response).

Although in the classical cinematogram stimuli (Braddick, 1974) the major difficulty is in resolving the correspondence problem, here the correspondence problem was minimized due to the low density of the dots. It is important to note that the cylinder could not be detected perceptually in each of the single frames (e.g. by detecting different density of the dots around its vertical borders), and that, in orthographic projection, the direction of rotation of the cylinder is ambiguous.

RESULTS

The rotational velocity of the cylinder was measured in terms of the angular separation between successive frames. The performance (percentage of correct responses) of two naive observers (with corrected-to-normal vision) as a function of the angle of rotation is given in Fig. 2. The actual 2-D displacement of the dots on the screen, measured in visual angle, was used to estimate the percentage of short-range moving dots (where the displacement was less than 20 min arc, which is an average upper limit for SR displacement for the eccentricities we used; see Baker & Braddick, 1985) among all moving dots of the cylinder. This percentage is denoted by the broken line in Fig. 2. It can be seen that the detection performance deteriorates with an increase in angular separation between the frames. At the level of 6 deg of rotation

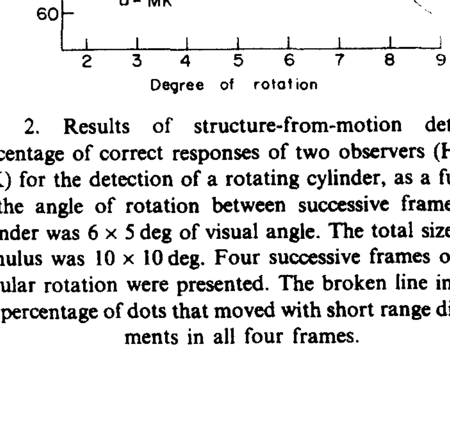

Fig. 2. Results of structure-from-motion detection. Percentage of correct responses of two observers (HZ and MK) for the detection of a rotating cylinder, as a function of the angle of rotation between successive frames. The cylinder was 6 × 5 deg of visual angle. The total size of the stimulus was 10 × 10 deg. Four successive frames of equal angular rotation were presented. The broken line indicates the percentage of dots that moved with short range displacements in all four frames.

between each frame, detection rapidly falls off, reaching 75% correct responses at 8 deg of rotation. It can be observed that the percentage of the population of SR moving dots decreases in a similar manner. Note that the total exposure time of the stimulus in our experiments is much shorter than that used by Petersik (1980) and Prazdny (1986).

To make sure that the results were not merely the consequence of incorrect correspondence caused by the large displacements, the experiments were repeated using a smaller number of dots spread in the same array. The results were not changed by this decrease in density.

In a following experiment, all short-range information was removed from the rotating cylinder. Because of the 2-D projection, most SR moving dots appeared near the vertical boundaries of the rectangular projection of the cylinder. The two sides of the projection of the cylinder were therefore removed, and the resulting space that was left was filled by adding background-like moving dots. The resulting percept of the cylinder was as if its sides were covered up with an opaque occluder. In most of the presentations the velocity of the background dots was kept near the average of the velocities of the cylinder dots. Another, higher, average velocity was also tested, which was the average of the projected part of the cylinder, (after removing the two low-velocities regions). Results of this experiment showed that about twice as many frames were necessary to perceive a coherent motion of the rotating cylinder. We do not think that the longer time required is merely the result of dot numerousity (i.e. that the number of dots belonging to the cylinder was reduced), since dot numerousity has little effect on SFM performance in such displays (Petersik, 1980). Rather, it appears that our SFM detection task under LR conditions may require a longer time to integrate together local information from different parts of the display.

DISCUSSION

The results obtained here are in agreement with recently published results by Todd et al. (1988) who determined the different combinations of display parameters that elicit a perceptually compelling impression of a solid object rotating rigidly in 3-D space. They found that the rotating cylinder is perceived as more rigid when moved in the SR. In their experiment all the moving dots belonged to the cylinder, while in ours noisy background dots were presented so that the relevant dots had to be segregated from the background.

The conclusion we propose is that SR motion information is necessary for fast and efficient detection of SFM. Under LR conditions the task is still possible, but not with brief presentations. SFM experiments under LR conditions have been described in the past (Petersik, 1980; Prazdny, 1986), but in these studies the presentation time was longer, and the percept was also aided by additional help from either explicit edges, or perspective depth cues. Our results here are consistent with earlier findings showing that LR motion detection is attentive and serial (Dick et al., 1987) and rules out the possibility that SFM has direct parallel access to LR motion information. It has been suggested in the past, however, that attentional resources can sometimes be distributed over the visual field, but with low resolving power (Bergen & Julesz, 1983; Eriksen & St James, 1986). Thus it is still possible that, in the absence of SR information, SFM may be computed in parallel, but with lower resolution.

REFERENCES

Anstis, S. M. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society, London B, 290,* 153–168.

Baker, C. L. & Braddick, O. J. (1985). Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Research, 25,* 803–812.

Bergen, J. R. & Julesz, B. (1983). Rapid discrimination of visual patterns. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13,* 857-863.

Braddick, O. J. (1974). A short-range process in apparent motion. *Vision Research, 14,* 519–527.

Braunstein, M. (1976). *Depth perception through motion.* New York: Academic Press.

Dick, M., Ullman, S. & Sagi, D. (1987). Parallel and serial processes in motion detection. *Science, 237,* 400–402.

Eriksen, C. W. & St James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics, 40,* 225–240.

Green, B. F. (1961). Figural coherence in the kinetic depth effect. *Journal of Experimental Psychology, 62,* 272–282.

Mather, G. (1989). Early motion processes and the kinetic depth effect. *The Quarterly Journal of Experimental Psychology, 41A(1),* 183–198.

Petersik, J. P. (1980). The effects of spatial and temporal factors on the perception of stroboscopic rotation simulations. *Perception, 9,* 271–283.

Petersik, J. P. (1989). The two-process distinction in apparent motion. *Psychological Bulletin, 106(1),* 107–127.

Prazdny, K. (1986). Three dimensional structure from long-range apparent motion. *Perception, 15,* 619–625.

Todd, J. T., Akerstrom, R. A., Reichel, F. D. & Hayes, W. (1988). Apparent rotation in three-dimensional space: Effects of temporal, spatial, and structural factors. *Perception and Psychophysics*, *43*, 179–188.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, Mass.: MIT Press.

Wallach, H. & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, *45*, 205–217.