CORRESPONDENCE IN APPARENT MOTION: DEFINING THE HEURISTICS

MARC GREEN

PSYCHOLOGY DEPARTMENT
YORK UNIVERSITY
NORTH YORK, ONTARIO M3J 1P3

ABSTRACT

Correspondence matching in apparent motion is based on two heuristics: match images if they 1) have a similar form and 2) are in close proximity. Psychophysical experiments are used to define these heuristics. Observers judged motion path between images in a competition paradigm. Results showed that the tokens used in form matching are spatial frequency and orientation. Further, proximity is defined in a 3-D spatial reconstruction rather than 2-D retinal coordinates. A possible representation for the computation of correspondence is a multidimensional detector space, with dimensions including spatial frequency, orientation, X, Y and Z (or disparity) coordinates.

INTRODUCTION

A remarkable property of biological vision systems is the ability to deduce that two images, seen at different places and/or times, represent the same physical object. The advantages of this property are nicely exemplified in the phenomenon of apparent motion: when viewing a series of static pictures, or "frames", each object in one frame moves to the location of the corresponding object in the subsequent frame. Coherent motion is perceived only if the visual system, after considering successive frames, can properly match images corresponding to the same object. This "correspondence problem" is presumably solved by application of heuristics to provide a "preference metric" (13) which evaluates the affinity between potential matches. Preference metrics can be derived by two general classes of heuristic: 1) match images of similar form and 2) match images with the greatest spatial proximity. At first glance, this recipe for matching seems simple enough, but real difficulties arise when implementation is attempted. These heuristics need to be more precisely defined by answering the following questions. First, what form primitives are used as tokens in correspondence matching? Second, is proximity defined in two-dimensional retinal coordinates or in an internal, 3-D reconstruction of space. This question has important implications since use of a 3-D metric requires that a depth must be assigned each form token before matching can proceed. The studies described below are addressed to answer each of these questions.

THE FORM HEURISTIC

The notion that correspondence matching is based partly on form similarity has been around for a long time. However, it has proved suprisingly difficult to identify correspondence tokens since apparent motion tends to be independent of form similarity. Early studies (8, 14) found that when there was only one image in each frame, the apparent motion seen with two identical images was readily perceived with two different images. The first would deform gradually into the second with no loss of motion continuity. More recently experimenters (3,11) have used competition methods and have likewise concluded that form similarity plays no role in token matching.

Why has it proved so difficult to identify correspondence tokens? Two possible explanations come to mind. First, stimuli were either geometric forms, circles, squares, letters, etc. or alphabetic characters, which differ in high spatial frequency content, but are similar in low spatial frequencies. Both human psychophysical (1) and computer (9) experiments have resulted in the view that one representational stage in early visual processing is the activity in arrays of detectors which are sensitive to edges at different resolution. At each resolution level, the detectors are activated only by a narrow band of spatial frequencies. Geometric shapes and alphabetic characters would all stimulate similar populations of coarse, low resolutions detectors. If activity in detectors at different resolution were tokens, then there would be strong affinity between all such images. Second, previous investigators have used a flash technique which produced a luminance transient (i.e., a D. C. offset) accompanying the presentation of form. The luminance flux *per se* might be used as a token for matching. This seems plausible because it has been suggested (4) that the visual system contains two parallel sets of detectors for analyzing spatio-temporal luminance change. The detectors are modeled as difference of Gaussians (DOG's) with different temporal properties. If the inhibitory Gaussian is developed simultaneously with the excitatory, then the detector is "sustained" and is highly tuned to aspects of form such as spatial

frequency and orientation. If inhibition is delayed, the detector is "transient", responds to D. C. flux and shows little selectivity to form. Correspondence might be determined by matching patterns of activity in these transient detectors.

I tested these possibilities in a set of psychophysical experiments (5). The initial assumption was that correspondence matching is mediated by the activity of detectors tuned to edges of different resolution and orientation. My strategy involved patterns which would be much more selective in the populations of detectors that were being stimulated. To eliminate the problem of common low frequency components, I used Gaussian modulated sinusoids or "Gabor functions". The spatial frequency content of Gabor functions is narrow and easily controlled by varying the period of the sinusoid. Luminance changes were eliminated by insuring that the time and space-averaged luminance of the Gabors were equal to that of the background.

Stimuli were displayed on a Hitachi high resolution monitor driven by a Grinnell graphics system. The viewing area was 14 by 12 degrees and had a mean luminance of 65 cd/m2. When no stimuli were being displayed, the screen was uniform in luminance with the exception of a central cross-hair provided for fixation.

Targets were Gaussian modulated sinusoids, or "Gabor functions". These were created by calculating a sine-wave function, which varied around the mean luminance, and then multiplying the sinusoid by a circular Gaussian. The final product appeared as a circular patch of sine-wave about 1.7 degrees in diameter in which contrast was maximum at the center and decreased radially. Unless otherwise stated, phase of the sinusoid was 0 degrees with respect to the Gaussian function. This was necessary to insure that the space-averaged luminance of the Gabor would always be the same as that of the background. Contrast of the Gabor functions was determined by a matching procedure. The Gabor containing the highest central frequency, 10 c/deg, was set to 85% contrast. Physical contrast of all other Gabors was set to the same apparent contrast.

Experiments consisted of a series of trials in which the observer viewed a sequence of 4 frames. As shown in Figure 1, each frame contained four Gabors drawn on the circumference of an imaginary circle. Frames consisted of two pair of identical Gabors ("A" and "B"). In the experiments, A and B represented different values along the dimensions of spatial frequency or orientation. Figure 2 shows a picture of the actual display. In this case A and B differ in spatial frequency by 1.5 octaves. Frames 2 through 4 consisted of the same stimuli rotated by 45 degrees to new positions. Rotation changed only position, not orientation of the Gabors. The correspondence problem asks how the Gabors in frame 1 decide which Gabor in frame 2 is a proper match. Since the distance from A to B or another A is equal, there is no a priori reason for motion to be clockwise or counter-clockwise. If the difference between A and B can be used to determine correspondence, the A moves to A and B to B. Otherwise, direction should be ambiguous.

The observers' task in all experiments was to discriminate clockwise from counter-clockwise motion. On each trial, the sequence of 4 frames

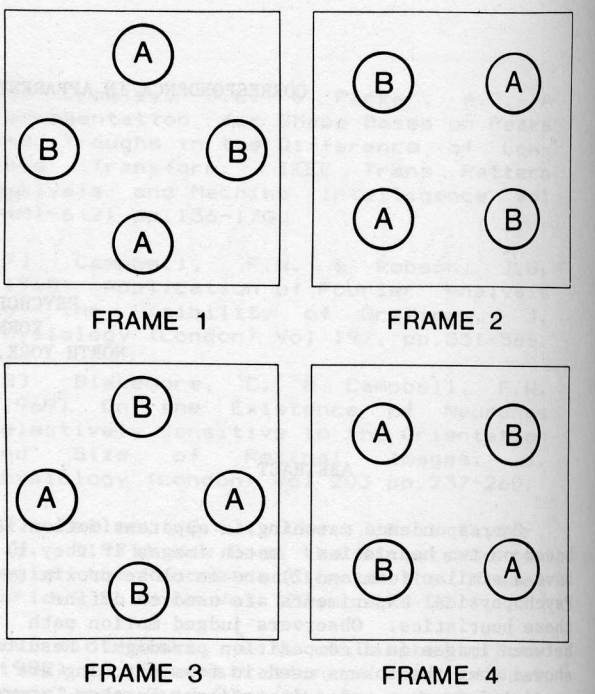

Figure 1

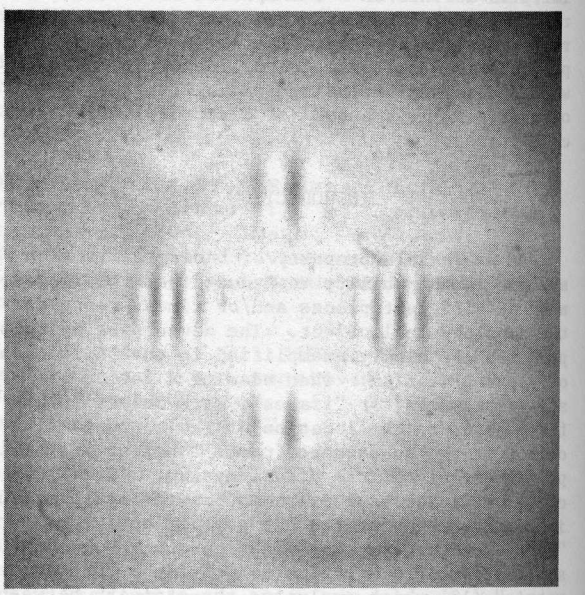

Figure 2

was shown twice in succession to produce rotation through 315 degrees. Frame duration was 84 msec (5 sweeps of the raster) and the interstimulus interval (ISI) between frames, during which only the uniform field was visible, was 17 or 50 msec. The only reason for choosing these time intervals was that they produced clear motion. The results reported below were robust and did not depend critically on any particular frame duration or ISI.

In the first set of experiments, A and B Gabors of different spatial frequencies. The left panel of Figure 3 shows the results obtained when A was fixed at 1.7 c/deg and the distance between

the centers of similar Gabors was 5.4 degrees. The actual distance between Gabors in successive frames was 2.3 degrees. This meant that there was no overlap in the position of a Gabor from one frame to the next. Ability to perceive direction of motion was at chance levels when A and B had the same value. As spatial frequency of B increased, discrimination between clockwise or counter-clockwise directions improved until performance was perfect. Observers reported that their ability to judge direction resulted from a coherent motion of the Gabors in a circular path. Data in the bottom panel show results obtained when spatial frequency of A was fixed at 5.0 c/deg. Clear spatial frequency tuning of the correspondence process is again evident for both observers. The tuning of the matching process is suprising sharp. I estimated that the curves fell to half-width/half-height in 0.5 to 1.0 octave, a value similar to that found for cells in the primary visual cortex and many other psychophysical experiments. I repeated the experiment with smaller diameter circles to produce different amounts of overlap between Gabor positions in successive frames. There was no evidence that matching was affected by whether or not overlap existed between successive frames (5).

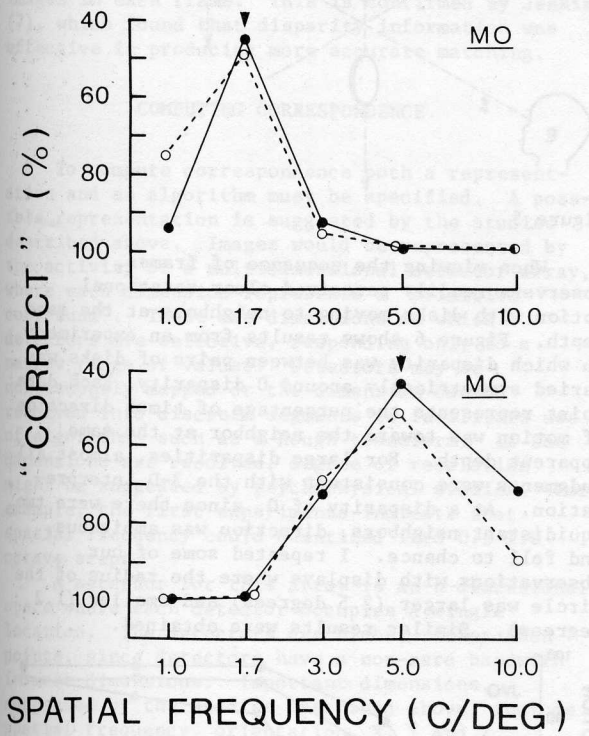

Figure 3

I earlier speculated that previous studies may have failed to find correspondence tokens because of the luminance flux which accompanied form presentation. To test this possibility, I repeated the basic experiment except that the background luminance was dark (actually 0.5 cd/m2) or half that of the Gabors (32.5 cd/m2). Observers failed to perceive strong coherent motion under any conditions.

I also investigated the possibility that orientation may be a token for correspondence. For this experiment spatial frequency was set at 3.0 c/deg and orientation difference between A and B varied. As shown in Figure 4, correspondence exhibits a clear tuning for orientation. Differences of 22.5 degrees from the A value produced almost perfect performance. Although performance was excellent with orientation as the correspondence token, observers agreed that the coherence of the motion produced by orientation, although clear enough to make a correct judgment, was never as smooth as for spatial frequency.

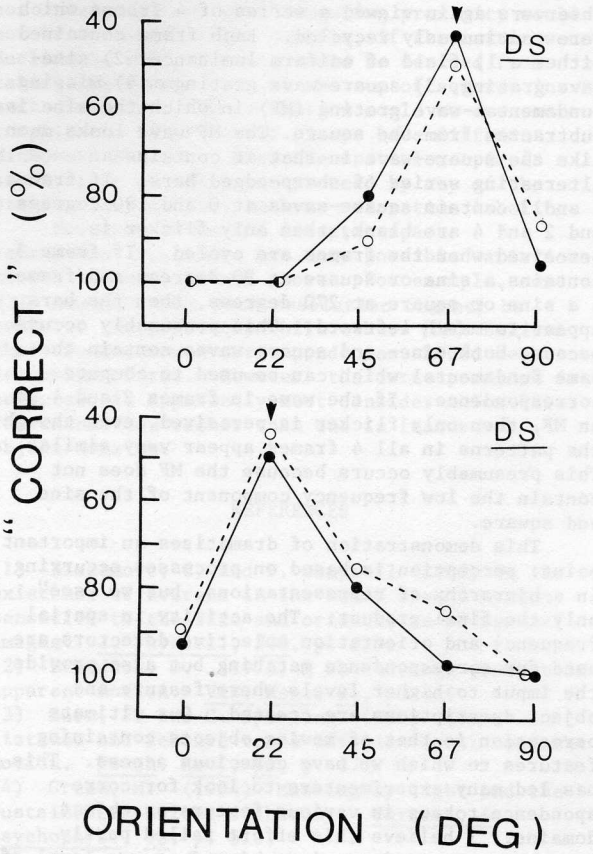

Figure 4

I concluded from these experiments that matching is based on similarity of spatial frequency and orientation. This contrasts greatly with the previous finding that form similarity seems to be important in apparent motion. My view is similar to that of Ullman (13), who has suggested that the many failures to uncover form tokens occurred because experimenters used relatively complex images. Shapes such as alphabetic characters consist of numerous tokens, so that any two images will usually have some tokens that match. By using simpler stimuli, isolated line segments, he found that orientation was an important token in matching. My conclusion differs slightly in that it is not the oriented lines that are important but rather the activity in populations of oriented, narrowband detectors. My analysis further differs from that of Ullman since I suggest that luminance transients play a large role in correspondence. The existance of transients may explain why orientation is not always found (3) to be a token, even with single

lines.

Some (2) believe that there are two mechanisms of motion correspondence, the "short-range" and "long-range" systems. Ullman also concluded that orientation was a token only when matching spanned very small spatial separations and stimulated short-range mechanisms. The step sizes used in my experiments were relatively large and likely activate long-range mechanisms. An additional demonstration shows that spatial frequency is also a token in short-range motion. Observers again viewed a series of 4 frames which were continuously recycled. Each frame contained either a 1) field of uniform luminance, 2) sine-wave grating, 3) square-wave grating or 4) missing fundamental-wave grating (MF) in which the sine is subtracted from the square. The MF wave looks much like the square-wave in that it contains an alternating series of sharp-edged bars. If frames 1 and 3 contain square-waves at 0 and 180 degrees and 2 and 4 are blank, then only flicker is perceived when the frames are cycled. If frame 3 contains a sine or square at 90 degrees and frame 4 a sine or square at 270 degrees, then the bars appear to march leftward. This presumably occurs because both sine- and square-waves contain the same fundamental which can be used to compute correspondence. If the wave in frames 2 and 4 is an MF, then only flicker is perceived, even though the patterns in all 4 frames appear very similar. This presumably occurs because the MF does not contain the low frequency component of the sine and square.

This demonstration of dramatizes an important point: perception is based on processes occurring in a hierarchy of representations, but we "see" only the final product. The activity in spatial frequency and orientation selective detectors are used for correspondence matching but also provide the imput to higher levels where feature and object descriptions are created. Our ultimate perception is that of moving objects containing features to which we have conscious access. This has led many experimenters to look for correspondence tokens in various feature or object domains. I believe this effort failed partly because correspondence is achieved at one level of representation while the features and objects were constructed at another. Too many experimenters employ images which indiscriminantly active detectors operating at low levels of representation. This is bound to obscure analysis of visual processing. The visual scientist's version of Occam's razor is that phenomena ought to be explained at the lowest possible level.

THE PROXIMITY HEURISTIC

The second unresolved question is whether matching employs two or three dimensional proximities. Initial studies (12) suggest that matching is based on 2-D proximities. In these studies, linear perspective was used to produce depth separation. I reexamined this conclusion using another depth cue, disparity (6).

The display was similar to that described above, except that each frame consisted of random dot stereograms, with A and B being disk-shaped submatrices of different disparity. Each disk was 1.3 degrees in diameter and lay on an imaginary circle with a 1.8 degree diameter. Figure 5 shows how the display appeared to the observers. The pairs of disks seemed to float in front of the background at different depths. A small red square of 0 disparity provided at fixation point. Observers viewed a series of 8 such frames in which the disks' positions were rotated by 45 degree steps. If correspondence matching is based on 2-D proximity in the XY plane, then direction of rotation is ambiguous: frame 2 contains two possible matches equidistant from each object in frame 1. If 3-D proximity is used as the distance metric, then objects will appear to move to the neighbor in the same depth plane and therefore closer in 3-D space.

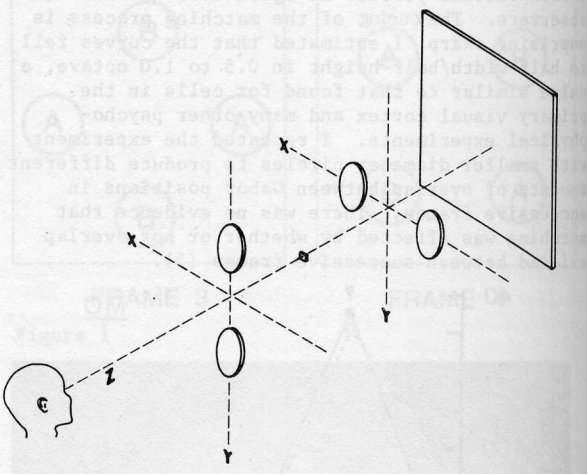

Figure 5

When viewing the sequence of frames, observers readily perceived clear rotational motion with disks moving to neighbors at the same depth. Figure 6 shows results from an experiment in which disparity was between pairs of disks was varied symmetrically around 0 disparity. Each data point represents the percentage of times direction of motion was toward the neighbor at the same apparent depth. For large disparities, almost all judgments were consistent with the 3-D interpretation. At a disparity of 0, since there were two equidistant neighbors, direction was ambiguous, and fell to chance. I repeated some of our observations with displays where the radius of the circle was larger (2.5 degrees) and smaller (1.2 degrees). Similar results were obtained.

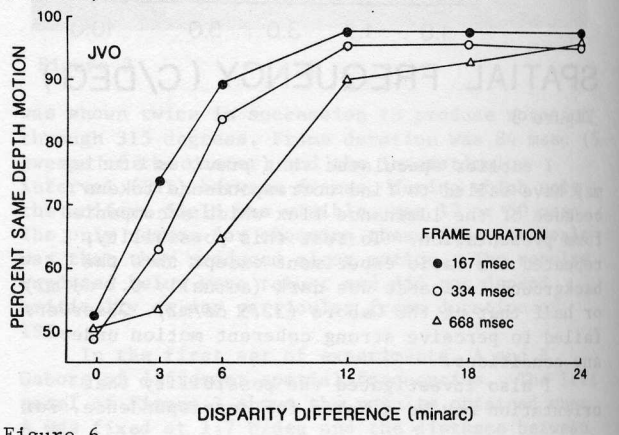

Figure 6

In the experiment described above, images had no monocular cues. We also were forced to use long durations since the disks would dissolve into the background if the motion were too fast.
We repeated the experiment with the light squares of the disks darkened so that the observers saw gray blobs floating against the background. This produced monocular cues and allowed us to use a higher frame rate, 84 msec per frame. The new display also produced a compelling circular motion with disks moving toward neighbors in the same depth plane, and observers were almost 100% correct in discriminating the motion path. To be sure that the monocular cues did not contribute to ability to judge direction, observers attempted the task with one eye occluded. Results showed the observers performing at chance.

Our results are consistent with the view that correspondence matching utilizes 3-D proximities. In a subsequent experiment, we further demonstrated that distance in the X, Y and Z planes can be traded off so that objects will appear to move in depth when the nearest neighbor lies at a different disparity. Based on our evidence, it might be expected that correspondence matching by computer would be enhanced by assignment of depth/disparity to images in each frame. This is confirmed by Jenkin (7), which found that disparity information was effective in producing more accurate matching.

COMPUTING CORRESPONDENCE

To compute correspondence both a representation and an algorithm must be specified. A possible representation is suggested by the studies described above. Images would be represented by the activity of a multidimensional detector array, where each dimension represents a "primitive continuum". These are dimensions on which detectors are selective, responding only to a narrow range of values. Detectors may be continuously mapped or the dimension can be resolved into discrete segments to facilitate use of algorithms such as a Hough transform. If the dimensions are resolved, degree of resolution might be suggested by psychophysical studies. For example, the first experiments indicate that spatial frequency could quantized into 0.5-1.0 octave steps.

A metaphor for this array is an n-dimensional space where each detector occupies a single location. The detectors are blobs, rather than points, since detectors have a non-zero bandwith in most dimensions. Important dimensions, suggested by the studies described above, include spatial frequency, orientation, X, Y and Z coordinates (or possibly disparity, depending on whether other depth cues are employed). Sustained and transient detectors would have to be separately represented. There may be other important dimensions as well. For example, I have been examining whether color might be added to this list, but the results so far have been ambiguous.

To perform correspondence matching the activity of points at time t1 is compared to activity of points at t2. A correspondence algorithm might try to match up similarly tuned detectors, i.e., nearest neighbors in the detector space. The preference metric then becomes an index of proximity in the space and would be calculated by summing the (weighted?) proximities in each of the n dimensions. To accomplish this, the metric of the detector space must be found. In the simplest case, dimensions are independent and the metric is "city-block". Proximity is then simply a sum of the distances in each dimension. However, it is more likely that the dimensions are not independent so that computing proximity would not be so simple. For example, if the metric were Euclidean, then distance would be derived from taking the square-root of the sum of squares in each dimension. The situation could be even more complicated if different planes have different Minkowski metrics. Once the spatial metric is known, then matching can proceed by any of the standard algorithms, such as relaxation labeling.

However, given the detector space representation, there are many possible variations to the scheme outlined above. For example, some (10) suggest that each resolution channel be computed independently while others (11a) believe that cross-channel correspondences should be determined first. However the correspondence is computed, it apparently must consider a low level representation rather than one in feature or object domains.

REFERENCES

(1) Blakemore, C. and F. Campbell (1969) On the existence of neurons in the human visual system sensitive to the size and orientation of retinal images. J. Physiol., 203, p. 237.
(2) Braddick, O. (1974) A short-range process in apparent motion. Vis. Res., 14, p. 519.
(3) Burt, P. and G. Sperling (1981) Time, distance and feature trade-offs in visual apparent motion. Psych. Rev., 88, p. 171.
(4) Green, M. (1984) Masking by light and the sustained-transient dichotomy. Percept. & Psychophys., 35, p. 519.
(5) Green, M. (1986) What determines correspondence in apparent motion? Vis. Res., in press.
(6) Green, M. and J. Odom (1986) Correspondence matching in apparent motion: Evidence for 3-D internal representation. To be published.
(7) Jenkin, M. (1983) Tracking three dimensional moving light displays. Proc. ACM Interdisc. Work. Mot., p. 66.
(8) Kolers, P. and J. Pomerantz (1971) Figural changes in apparent motion. J. Exp. Psych., 87, p. 99.
(9) Marr, D. (1982) Vision.
(10) Marr, D. and E. Hildreth (1979) The theory of edge detection. Tech. Report from M. I. T.
(11a) Mayhew, J. and J. Frisby (1981) Psychophysical and computational studies towards a theory of human stereopsis. Artif. Intell., 17, p. 349.
(11) Navon, D. (1976) Irrelevance of figural identity for resolving ambiguities in apparent motion. J. Exp. Psych. HP & P, 2, 130.
(12) Ullman, S. (1978) Two dimensionality of the correspondence process in apparent motion.

Perception, 7, p. 683.
(13) Ullman, S. (1980) The effect of similarity between line segments on the correspondence strength of apparent motion. Perception, 9, p. 617.
(14) Wertheimer, M. (1912) Experimentalle Studien uber das Sehen von Bewegung. Z. Psychol., 61, p. 161.