Vision Research, 40(6):695-704, 2000.

Effect of 3D Structure on Motion Segmentation

MARY J. BRAVO, [1] HANY FARID, [2]

A smooth surface imaged on the retina produces a smooth flow field. Thus, the visual system may group regions of smoothly varying flow to segment surfaces. We tested this idea by having observers perform a segmentation task on several stimuli that differed in their 3D interpretations but were all matched in the smoothness of their 2D flow fields. Performance varied across conditions with the best performance occurring when the stimulus simulated a rigid plane. This result suggests that while observers may use deviations from smoothness to segment a broad class of motion stimuli, they use a more precise strategy to segment stimuli with a familiar 3D interpretation.

2D-Motion 3D-Motion Segmentation Grouping

[1]Department of Psychology, Rutgers University, Camden NJ 08102.
[2]Perceptual Science Group (NE20-444), MIT, Cambridge MA 01239.

1 Introduction

Objects moving in an observer's field of view produce a flow field on the retina. Gibson was one of the first to emphasize that this flow field provides a rich source of information about the environment and our movements within it [11]. More recently, Koenderink and others have shown precisely how the structure of the world is reflected in various properties of the flow field [15]. In theory then, we can acquire a great deal of useful information about the 3D world from computations performed directly on this 2D motion pattern. This information includes the orientation of surfaces [15, 10], our direction of heading [11, 29], and our time to contact an approaching object [16].

The most fundamental relationship between the structure of the world and the structure of the flow field is smoothness. Simply put, smooth surfaces produce smooth flow [17, 15] . This relationship, embodied as a smoothness constraint, is fundamental to models of the recovery of the 2D flow field from ambiguous or noisy motion signals [13, 31, 30]. Smoothness is also used either explicitly or implicitly in models of surface segmentation [3]. A corollary to the observation that smooth surfaces produce smooth flow is that a flow discontinuity indicates a surface discontinuity. This suggests two complementary strategies for segmenting a dynamic image: the aggregation of similar velocities and the segregation of dissimilar velocities. Most of the work on human motion segmentation has focused on the later. The idea that the visual system should have specialized detectors for velocity discontinuities has been around for some time [20] and there is both physiological and psychophysical evidence that such detectors exist [1, 23, 27]. Because these detectors compute spatial derivatives of the flow field they are effectively measuring deviations from smoothness.

The purpose of the work reported here was to test the idea that performance on a motion segmentation task can be predicted by considering only the smoothness of the 2D flow field. If so, stimuli that are matched in the smoothness of their flow fields should produce similar performance in a segmentation task regardless of the 3D percepts they evoke. We tested this prediction by presenting observers with a set of stimuli that had different 3D interpretations but flow fields that were matched for smoothness.

[3] We do not think of flow field recovery and segmentation as discrete steps in motion processing. Research has tended to focus on one problem or the other, but it is widely assumed that these two processes are intimately related.

Essential to this experiment was our animation method. Studies of the 3D perception of structure from motion generally use the standard rendering approach of defining a 3D model, projecting it onto the image plane, moving the model and then projecting it again. This technique allows the experimenter to have precise control over the 3D surface and 3D motion being simulated, but does not permit direct control, or even direct knowledge of, the flow field. In contrast, studies of 2D motion perception vary the pattern of image velocities without concern for whether the stimulus simulates a 3D object. Our experiment required the manipulation of the 2D flow field produced by a 3D object. So, to animate our stimuli, we used flow equations that specify for a given 3D surface and 3D motion how the image velocity varies as a function of image location and time. Simple transformations of this flow field allowed us to generate control stimuli that were identical with respect to a particular 2D property (here smoothness) but differed in their 3D interpretation.

2 Methods

2.1 Motion

Our test condition simulated the rigid rotation of two planar surfaces with different slants. One surface was large and spanned the display, the second was much smaller and was embedded in the first. The stimulus resembled a large rotating wall that contained a small open door. This stimulus rotated about a vertical or horizontal axis that was located on the larger plane and intersected the line of sight. Figure 1 shows the geometry of the simulated stimulus viewed from the front and from the top. The background and target planes rotated rigidly, so the angle between them was constant within a trial and was either 10, 20 or 30 degrees.

We animated these stimuli using flow equations derived for a rigidly rotating plane viewed under perspective projection. Below are the equations for the background and target planes undergoing a rotation about the same vertical axis. The flow field of the background plane was:

$$V_x(x,y,t) = \tan(\omega t)\omega x + \frac{\omega}{f}x^2 \quad (1)$$

$$V_y(x,y,t) = \frac{\omega}{f}xy, \quad (2)$$

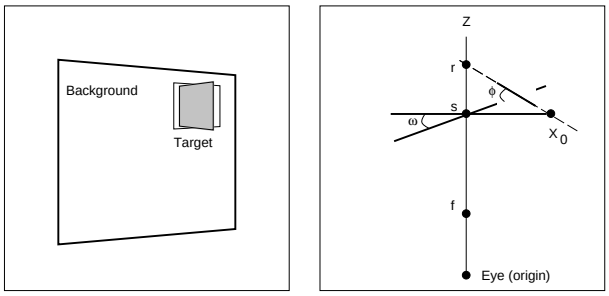

Figure 1: The stimulus simulated a large rotating plane (background) with a small open door (target). This stimulus rotated rigidly about a vertical or horizontal axis. Shown are front and top views. The eye is at the origin, the computer screen is at f, and the simulated plane is at s. The variables f, s, and r represent distances from the origin along the vertical Z axis, X_0 is the distance at which the target plane intersects the horizontal axis, ω is the angle of the background plane relative to horizontal, and ϕ is the fixed angle between the background and target planes.

and the flow field of the target plane was:

$$V_x(x,y,t) = \omega f\left(1-\frac{s}{r(t)}\right) + \frac{s}{r(t)}\tan(\omega t+\phi)\omega x + \frac{\omega}{f}x^2 \quad (3)$$

$$V_y(x,y,t) = \frac{\omega}{f}xy, \quad (4)$$

where the variables are defined in Figure 1, and $r(t) = s - \frac{X_o \sin(\phi)}{\cos(\omega t+\phi)}$.

With these equations we can see how changing the angle between the target and background planes (ϕ) changes the flow field. First, note that the V_y component of the flow field is the same for the target and background planes. Also note that for both planes the V_x component of the flow field is a function of the horizontal position in the image (x) and time (t) but not of the vertical position in the image (y). Thus, changing the angle between the two simulated planes changes only the way V_x varies with x and t. And as Figure 2 shows, the relationship between the V_x component of the flow field and the angle between the two planes is a fairly simple one. The top row shows top views for the simulated stimuli at different points in time. Each graph shows target planes at three different angles with the background plane, but only one target plane was presented on a trial. The middle row shows how the V_x component of the

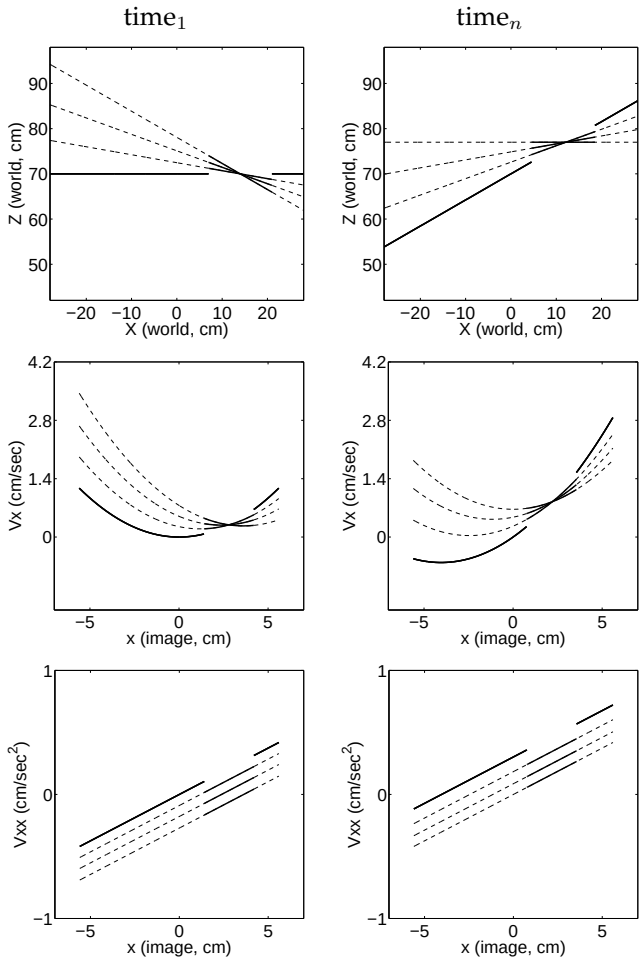

Figure 2: These graphs show the effect of varying the angle between the target and background planes on the smoothness of the flow field. The two columns correspond to different points in time. The top row depicts the 3D stimulus, the middle row depicts the horizontal component of the instantaneous flow field, V_x, and the bottom row depicts the partial spatial derivative of this velocity component, V_{xx}. For purposes of comparison, three targets planes are shown in each graph, but only one of these planes would have appeared in the stimulus. Note that as the angle between the target and distractor planes increases the difference between V_{xx} increases proportionally.

flow field varies with x. The intersection of the target and background flow fields corresponds to the intersection of the planes. From this plot it is clear that the average velocity of the target does not deviate in an appreciable way from the background. Instead, as the bottom row shows, the target deviates from the background in its velocity gradient. These plots indicate how the horizontal component of the velocity gradient, $V_{xx} = \partial V_x / \partial x$, varies with x for the target and background. As the angle between the target and background planes increases, the difference between V_{xx} increases proportionally. Across the range of stimulus rotations and target locations that we used, this relationship is approximately constant. So for this set of stimuli, increasing the angle between the target and background planes causes a proportional increase in the amount by which the target's velocity gradient deviates from that of the background.

The flow equations are clearly useful for gaining insight into the relationship between the distal stimulus, the moving 3D structure, and the proximal stimulus, the 2D flow field. But for our purposes, there is an even greater advantage to having these equations - they allow us to create control stimuli. Simple transformations of these equations can produce new stimuli that have exactly the same smoothness as the original flow field. And this allows us to test the idea that the smoothness of the flow field governs segmentation.

To generate our control stimuli we changed uniformly the sign or direction of the velocity vectors without altering their magnitudes. These changes do not affect smoothness, but they do change the possible 3D interpretation of the stimulus. This is equivalent to reflecting each velocity vector about horizontal. For the first set of control stimuli, we negated the V_y component of the flow field. For the second set of control stimuli we negated the V_x component and then swapped it with the V_y component. This is equivalent to rotating each velocity vector by ninety degrees. Examples of the planar and control flow fields are shown in Figure 3. We should emphasize that these transformations preserve speed, and, although they change directions, they preserve the direction gradients.

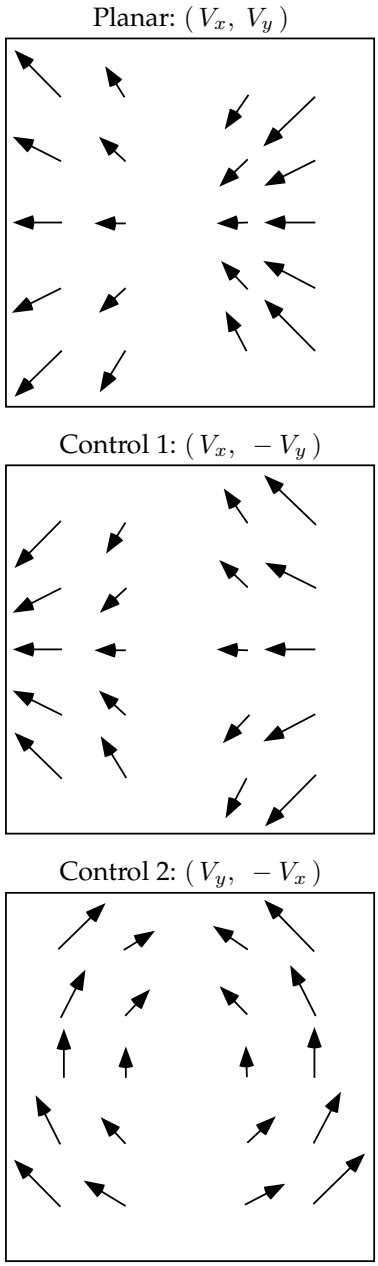

Figure 3: The top panel depicts an instantaneous flow field for the background plane generated according to Figure 1 and Equations (1) and (2). In the next two panels are the flow fields for the background plane in the control stimuli. These control stimuli were created by simple transformations of the planar flow field. These transformations preserve the 2D smoothness of the flow field but change the possible 3D interpretation.

2.2 Smoothness

We have stated repeatedly that our stimuli are matched in smoothness. But since we do not know how the visual system measures smoothness, this claim requires some justification. Here we attempt to show that for a

broad class of smoothness measures our stimuli would be equivalent. We start with a generic definition of smoothness: the integral over space of the sum of the squared partial spatial derivatives of the velocity field (Equations (1)-(4)):

$$\sum_{i=1}^{\infty}\left[\int_x dx\int_y dy\left(\frac{\partial^i V(x,y)}{\partial x^i}\right)^2+\left(\frac{\partial^i V(x,y)}{\partial y^i}\right)^2\right] \quad (5)$$

Our stimuli are matched in smoothness under this definition. There are, however, several reasons for thinking that this definition may not apply to human vision.

The most obvious shortcoming of this definition is that it ignores Weber's law. McKee showed that the speed difference needed for reliable discrimination depends on the base speed [18]. By taking derivatives we discard information about base speed and so we effectively violate Weber's law. This objection to the definition of smoothness does not apply to our control stimuli, however, since we created the controls by rotating the velocity vectors without changing their length. As a result, base speed was unchanged, and so our stimuli are matched in smoothness even when Weber's law is taken into consideration. A second shortcoming of the definition is that it is a global measure; the spatial derivatives are integrated across the entire display. Our visual system, on the other hand, probably measures smoothness at multiple spatial scales. But while spatial scale may be a problem for the definition of smoothness, it is not a problem for our stimuli since the conditions were matched in smoothness at all spatial scales. A third shortcoming of the definition is that it is insensitive to changes over time. The projected motions of surfaces are constrained to be smooth in time and space, and so a complete definition of smoothness must include temporal derivatives. But once more, this objection applies only to the definition, and not to our stimuli which had exactly the same temporal derivatives.

We must note one other important characteristic of this definition, and that is that it weights all velocity gradients equally. So for example, it treats compressive gradients (e.g., $\partial V_x/\partial x$) the same way as shearing gradients (e.g., $\partial V_x/\partial y$). The definition is also insensitive to the sign of these gradients. If the visual system has a significant anisotropy in its measurement of motion gradients, then this definition would be wrong and, more importantly, our stimuli would not be matched. We return to this potential shortcoming in the discussion section.

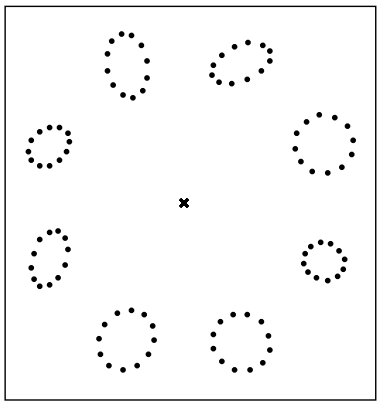

Figure 4: A single frame of our stimuli.

2.3 Texture

On the first frame of each animation sequence, eight ellipses were arranged around an imaginary circle with a radius of 11 degrees of visual angle centered on a fixation mark (Figure 4). Seven of these ellipses were animated using the flow equations for the background plane (Equations (1) and (2)). The flow equations for the target plane animated the eighth ellipse (Equations (3) and (4)). The observers task was to locate this eighth ellipse. The aspect ratio of each ellipse was selected randomly from a range of 1 to 2 (2.5 to 5.0 degrees visual angle) and the long axis was randomly oriented. We used a range of ellipse aspect ratios and orientations to obscure any shape cue that might be used to recover stimulus slant. A control experiment confirms that shape was not an important factor in performance (Section 4). Each ellipse was defined by 12 dots evenly spaced along its perimeter. The white dots were 5 arc min wide and were clearly visible on the black background.

2.4 Procedure

The stimuli were generated in MATLAB. They were displayed at 72 Hz on a computer monitor using routines from the Toolboxes developed by [2] and [22]. Subjects viewed the display monocularly through a cardboard aperture that obscured everything in their view except for the stimulus. A chin rest was used to help the subjects maintain a constant viewing distance of 20 cm [4].

[4] This viewing distance was slightly longer than that used in the flow field calculations. The reason for this discrepancy is that we were restricted in the size of the movie we could display but we needed a large field of view so our stimuli would have an appreciа-

Six subjects were recruited from the undergraduate population at the University of Pennsylvania. One was a practiced observer, the rest had never before participated in a psychophysical experiment. The initial session with each subject was devoted to practice. Subjects first learned to associate each of the eight ellipse positions with a key on the keyboard. Subjects then ran 36 trials on each of the three conditions. We used a large slant deviation and rotation angle (both were set to 30 degrees) so that the target would be fairly easy to detect. Larger angles would have made the target even more salient, but would also have produced a noticeable shape difference between the target ellipse and the background ellipses. During this practice phase, each stimulus was presented for 10 seconds and the subjects were permitted to move their eyes. Auditory feedback was provided.

For the experimental trials, the stimulus rotated back and forth through 30 degrees and two cycles of this motion were shown during the 2 second presentation interval. The experiment involved four 1-hour sessions. A session consisted of three blocks of 96 trials, one for each condition. The order of the conditions was randomized across sessions. Subjects were instructed to fixate a central cross during a trial and were given auditory feedback after incorrect responses.

The feedback was critical to this experiment since subjects were instructed only "to find the ellipse that does not belong with the rest". Subjects were told nothing about the expected 3D structure of the stimuli and had to learn, guided by feedback, what distinguished the target from the background. At the end of the practice session we asked the subjects to describe both what they saw and how they selected the target. Every subject reported seeing the planar condition as a flat surface rotating in depth and they reported selecting the ellipse "that stuck out". Descriptions of the control stimuli were more varied. Most saw control 1 as a curved surface that was slightly non-rigid and they again selected the ellipse that "stuck out". Two subjects saw this stimulus as flat and non-rigid and they selected the target based on its motion rather than its apparent depth. Control 2 appeared non-rigid to all of the subjects, and their reports varied from "something rubbery" to a "wiggling snaky thing". Most subjects saw this stimulus as having no depth, and they selected the ellipse that "moved differently from the rest".

ble amount of perspective. Without strong perspective, our transformations would have had a negligible effect on the flow field. We derived the flow field with a 14 cm viewing distance. However, we felt was that this viewing distance would be too uncomfortable for the subjects and so we used a longer distance in the experiment. This mismatch between the simulated and actual viewing distance seemed to have no effect on the apparent rigidity of the stimulus.

2.5 Task

This task of finding the odd ellipse was intended to tap into motion segmentation processes. However, in many ways the task resembles traditional visual search [26]. In a visual search experiment the subject must judge whether a particular target is present in a display containing a variable number of distractors. Logically, the task does not require that the distractors form a group or that the target segment from the distractors. Instead the visual system may simply look for activity in the detectors tuned to the target's distinguishing feature(s). As Bravo and Nakayama have shown, it is unlikely that segmentation is involved in visual search even when the target appears to "pop-out" from the distractors [6]. However, a feature detection strategy would not work well with our stimuli, because there was no simple feature or combination of features which defined the target. As Figure 2 shows, the motion of the background spanned a range of velocities which included the velocities of the target. What distinguished the target in our displays was that it did not fit the *pattern* of background motions. This claim is consistent with the subjective reports of our observers for the control conditions. As noted in the methods, observers said that they found the target in these displays by looking for the motion "that did not fit". The subjects reported using a different strategy for the planar condition. With these stimuli, subjects searched for the thing that stuck out. But it should be noted that sticking out is not a property of the target that exists in isolation from the background ellipses. In fact, when the target ellipse is presented by itself, it often does not appear to be rigid or to be moving in depth. It is only when several ellipses are distributed across the display that it is easy to discern the rigid rotation of the plane in depth. Thus we think that the strategy our subjects used to find the target is better characterized as segmentation than as feature detection.

3 Results and Discussion

We start with the data from the subject who had the best overall performance. The top graph in Figure 5 shows his data for the planar condition, that is the condition which simulated the rigid rotation of two planes. This graph shows the percentage of trials on

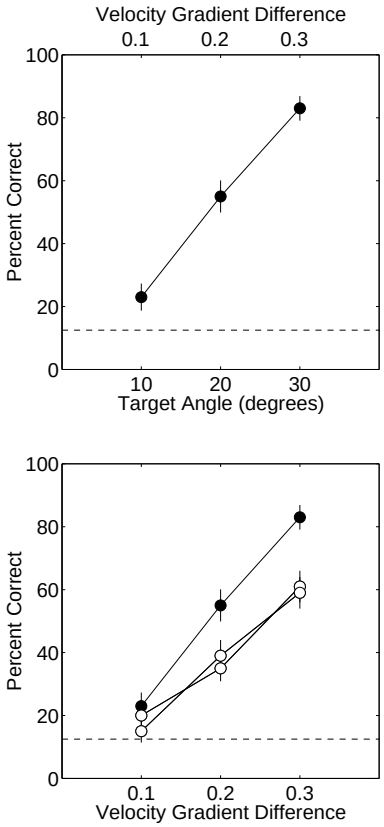

Figure 5: Shown in the top panel are results for the planar condition for one subject. The percentage of trials in which the target was correctly located is plotted against the angle between the target and background planes (bottom axis) and the difference in 2D velocity gradients (top axis). These data are replotted in the bottom panel along with the data from the control conditions (open circles).

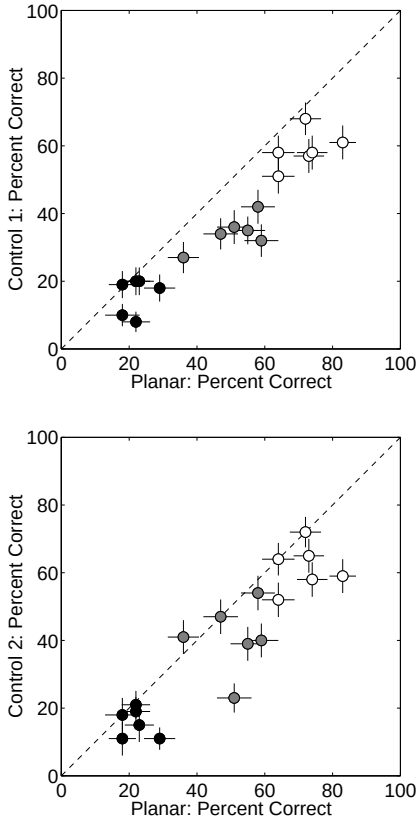

Figure 6: Results from all six subjects. Performance on the control conditions is plotted against performance on the planar condition. In both plots, the black, gray and white circles correspond to a 10°, 20° and 30° angle between the target and background planes.

which the subject correctly located the target plane plotted as a function of the angle between the target and background planes. Since the subject selected from eight locations, chance performance corresponds to 12.5%. Note that this subject performed above chance when the slant of the target and background differed by as little as 10 degrees. Not surprisingly, as the difference between the slant and target planes increased, performance improved.

Clearly this subject was sensitive to the difference in slants of the target and background planes. What we are interested in, though, is determining the basis of this sensitivity. The subject claimed that he perceived a 3D stimulus and that he selected the target based on its 3D appearance. Nonetheless, the information he used to locate the target is contained in the 2D flow field and it may be that the subject based his judgment simply on the deviation from smooth flow. As described in the methods section, the angle between the target and background planes was proportional to the difference between the velocity gradients of the target and background flow fields. If performance was based on detecting this deviation from smooth flow, then we would expect similar results for control stimuli that have a matching deviation. To make this comparison we first replot the data in Figure 5, this time using as our independent variable a 2D stimulus property (velocity gradient difference) rather than a 3D property (target angle). This allows

us to plot the data for the controls in the same graph since, by design, the control and planar stimuli were equivalent in terms of this 2D property. It is clear from the bottom graph that performance was not the same across these conditions. This subject was more accurate when locating the target in the rigid, planar stimulus than in the non-rigid, non-planar stimuli. This result indicates that the subject was not basing his judgment simply on the deviation from smoothness in the flow field. It suggests that motion segmentation is affected by the 3D interpretation of the stimulus.

Most of our observers showed a similar pattern of results: performance was generally better with the rigid, planar stimulus than with the controls. To facilitate the comparison of the planar and control conditions, we have plotted these data against one another in Figure 6. Each circle corresponds to the data from one observer and the color of the circle indicates the slant difference (10, 20 or 30 degrees). If observers had performed similarly on the planar and control conditions then the data would fall along the diagonal line in each plot. Most of the data fall below this line indicating that the observers generally performed better on the planar condition than on the control conditions. While the data are fairly neatly clustered for control 1, there is considerable variability in the data for control 2, with four points falling on or above the diagonal line. These points belong to two observers who performed the same on this control and the planar condition. Despite this intersubject variability, the overall pattern of results shows that performance in this task cannot be predicted solely from the smoothness of the flow field. These results suggest that stimuli with simple, familiar 3D structure have an advantage in motion segmentation.

4 Control Experiment

At the crux of our argument is the claim that these stimuli are matched in the smoothness of their flow fields. However, claiming that we have matched the flow fields that is *applied* to these stimuli is not the same as claiming that we have matched the flow fields that can be *measured* from these stimuli. For many stimuli, the discernible flow field differs from the true flow field, and this is especially true for stimuli containing extended contours. We used dot stimuli to avoid this "aperture problem", but it is still conceivable that motion processing occurs at multiple spatial scales and that at a coarse scale these dots might not be resolved. If these stimuli were blurred such that each ellipse appeared as a continuous form, this would surely "unmatch" our stimuli. In particular, control 2 contained large amounts of curl relative to the other conditions, and so a significant amount of the motion was directed along the contours of the ellipses. This motion would not be detected if the dots were not resolved. To ensure that our results were not critically dependent upon dot arrangement we repeated the experiment with jittered ellipses. In these stimuli the position of each dot was jittered by $\pm 30\%$ of its distance to the ellipse center of mass. Thus the dots formed irregular shapes lacking in smooth contours (Figure 7).

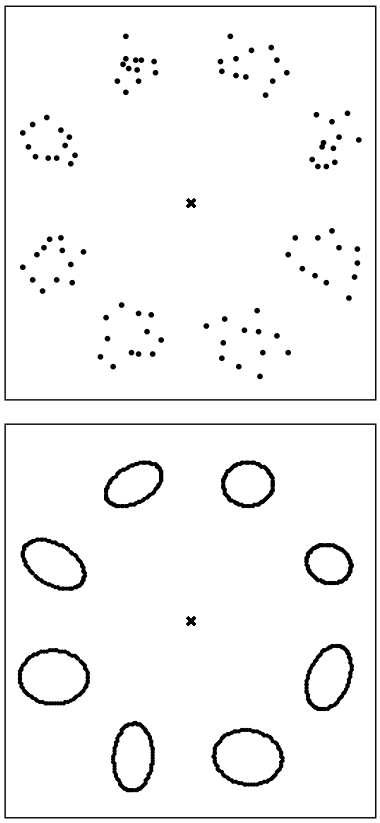

Figure 7: A single frame of our "jittered" and "continuous" control stimuli.

As a further control we also reran the experiment with ellipses that had continuous contours (Figure 7). If performance is based on motion, as we have assumed, then these stimuli should prove difficult for observers since the contours will obscure some components of the flow field. On the other hand, one

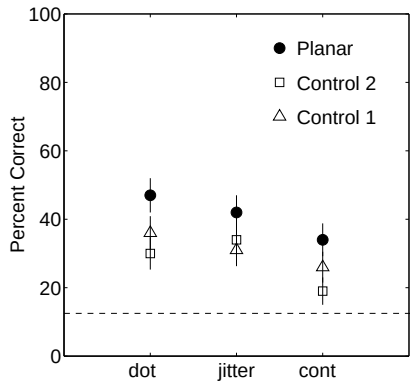

Figure 8: Shown is the percentage of trials on which one subject correctly located the target plane plotted for the original dotted ellipses, the jittered ellipses, and the continuous ellipses. The filled circles correspond to the planar condition and the gray and white circles correspond to the control conditions.

might argue that the regularity of the ellipses allows observers to base performance on a shape cue. If so, then it is the jittered ellipses that should present the greatest difficulty for observers. Thus by repeating the experiment with jittered and continuous ellipses we hoped to address the general concern that the pattern of results in the previous experiment was contingent upon the spatial characteristics of the display.

This experiment involved four naive subjects, two of whom had participated in the previous experiment. This experiment was identical to the previous experiment except that the difference in slant between the target and background planes was fixed at 25 degrees.

4.1 Results and Discussion

The pattern of results for this control experiment was consistent across the four observers and across the planar and control conditions. Performance was similar for the jittered and dotted ellipses, but worse for the continuous ellipses. Representative data from one observer are shown in Figure 8. Because performance with the continuous stimuli was consistently (although not always significantly) worse than with the other two stimuli, we conclude that observers did not base their judgment directly on changes in the ellipse shape but instead used the velocity field.

5 Discussion

There are two ways to describe our experiment. One focuses on the stimulus displayed on the computer monitor, the other on the observer's percept. Our stimulus was a dot texture that was animated by a 2D motion pattern specified by Equations (1) and (2). Across trials we made small local perturbations to this motion pattern and asked observers to locate these perturbations. A second description of our experiment is the one our observers gave. They reported seeing a rotating plane that had a small patch sticking out in depth. Their task was to find this protruding patch. As with all structure from motion (SFM) experiments, it is not immediately clear which of these descriptions is most relevant. There has been a long-standing controversy in the SFM literature over how to determine whether the 3D interpretation of a stimulus plays a role in any given task [24, 4, 3].

Our way of addressing this problem is to use stimuli that are matched in the relevant 2D property, but differ in their 3D interpretations. This strategy required using an animation technique that is not commonly used in SFM experiments. SFM stimuli are usually created by defining a 3D model, projecting it onto the image plane, moving the model and then projecting it again. The flow field is produced indirectly, but the experimenter knows that whatever it is, it will simulate a "real" object. For this experiment, it was critical that we control both the 2D and the 3D properties of our stimuli; that we simulate a real object and that we do so using flow equations that provide an analytic expression of the flow field.

The object that we simulated consisted of two planes that rotated rigidly together but had different slants. One plane spanned the display and served as a background, the other, smaller plane was embedded in this background and served as the target. We asked observers to locate this target plane and we used their ability to do so as a measure of segmentation. Since this stimulus was created using flow field equations, we could transform these equations to generate new flow fields that had the same 2D smoothness but different 3D interpretations. This allowed us to control the property of the flow field that is thought to be critical for motion segmentation. If the smoothness of the flow field can be used to predict motion segmentation, then stimuli that are equally smooth should produce equivalent performance. They did not; observers performed better with the rigid planar stimulus. We take this as evidence that the 3D structure of the stimulus plays some role in motion segmentation.

It might be possible to salvage the idea that motion segmentation is based solely on the smoothness of the flow field by arguing that we are using the wrong definition of smoothness. As we described in the methods section, a definition of perceptual smoothness would probably incorporate Webers' law for speed discrimination, temporal derivatives of the flow field, and multiple spatial scales at which smoothness is measured. But no matter how these factors are incorporated into a measure of smoothness, our stimuli will be matched. Of course, our stimuli are not identical, and so they cannot be matched for every conceivable definition of smoothness. In particular, our stimuli are not matched under a definition of smoothness that is significantly anisotropic. That is, if the perceived magnitude of a velocity gradient depends on either its sign or its direction, then these stimuli may not be equally smooth. There is some evidence that at detection threshold, the visual system has differential sensitivities to compressive and shearing gradients [21]. But it is not clear how to extend this result to the discrimination of suprathreshold 2D gradients. In any case, it is certainly possible to argue that our results are due to some combination of anisotropies in the measurement of velocity gradients. We cannot refute such an account, but we think that the simplest (and probably correct) explanation for our results is that the 3D properties of the stimulus play a role in motion segmentation.

If the smoothness of the 2D flow field is insufficient to predict performance on a segmentation task, does this mean that motion segmentation involves a 3D representation of the stimulus? Not necessarily, a purely image-based approach to segmentation may be biased for the flow fields that arise from familiar 3D surfaces and motions. In computer vision, such approaches are common. These approaches first make an assumption about the flow fields produced by surfaces in the world, and this assumption defines the type of parametric model that is used to describe the flow field. This assumption might be that the flow fields produced by surfaces are translations [7] or that these motion patterns are affine transformations [28]. If the flow field reflects the motions of multiple surfaces, then a complete description of the flow field will require multiple models. The end result is a characterization of the flow field in terms of a reasonably small number of models, with each model corresponding to a different surface. So, although this approach is image-based it can incorporate information about the motion patterns produced by familiar 3D surfaces.

There is already some evidence that the visual system uses such a model fitting approach for segmentation. Numerous experiments have demonstrated that the primate visual system is selectively sensitive to expansion and rotation patterns [25, 8, 9, 12, 19]. Recently, Bravo showed that observers can use these patterns for segmentation [5]. That is, she showed that observers presented with a global expansion or rotation pattern can locate a motion that is inconsistent with the pattern and that this judgment is not based on local velocity gradients.

Taken together with the present experiment, this research indicates that the visual system does not rely solely on deviations from smoothness for motion segmentation. While the earlier work of Bravo is readily explained in terms of internal models for familiar global flow patterns, it is less clear how to explain the present results. These results may reflect the existence of an, as yet, undiscovered global flow detector that is sensitive to the spatial and temporal pattern associated with a rotation in depth. Or the results may reflect a second stage of motion segmentation that operates on a 3D representation. After first organizing image motions using the criterion of 2D smoothness, the visual system may then attempt to recover a rigid 3D structure from each of the resulting groups. This process of building a 3D representation may require a further segmentation of the scene, as suggested by [14]. Additional experiments will be needed to characterize the motion segmentation processes that the visual system uses to complement a smoothness-based approach. The present results indicate that such processes exist.

Acknowledgments

The authors gratefuly acknowledge funding support from NIH Grant RO1 EY10424 and NSF Grant SBR 97-29015 for MJB and NIH Grant EY11005-04 and MURI Grant N00014-95-1-0699 for HF.

References

[1] O. Braddick. Segmentation versus integration in visual motion processing. *Trends in Neuroscience*, 16:263–268, 1993.

[2] D.H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:443–446, 1997.

[3] M. L. Braunstein. Structure from motion. In A.T. Smith and R.J. Snowden, editors, *Visual Detection*

of Motion, pages 367–393, San Diego, 1994. Academic Press.

[4] M.L. Braunstein and J.T. Todd. A distinction between artifacts and information. *Journal of Experimental Psychology: Human Perception and Performance*, 16:211–216, 1990.

[5] M.J. Bravo. A global process in motion segregation. *Vision Research*, 38:853–863, 1998.

[6] M.J. Bravo and K. Nakayama. The role of attention in different visual-search tasks. *Perception and Psychophysics*, 51:465–472, 1992.

[7] T.J. Darrell and E.P. Simoncelli. Separation of transparent motion into layers using velocity-tuned mechanisms. In *Investigative Opthalmology and Visual Science*, page 1052, Sarasota, FL, 1993.

[8] C.J. Duffy and R.H. Wurtz. Sensitivity of MST neurons to optic flow stimuli. I. a continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65:1329–1345, 1991.

[9] T.C. Freeman and M.G. Harris. Human sensitivity to expanding and rotating motion: effects of complementary masking and directional structure. *Vision Research*, 32:81–87, 1992.

[10] T.C. Freeman, M.G. Harris, and T.S. Meese. On the relationship between deformation and perceived surface slant. *Vision Research*, 36:317–322, 1996.

[11] J.J. Gibson. *The perception of the visual world*. Houghton Mifflin, Boston, 1950.

[12] M.S. Graziano, R.A. Andersen, and R.J. Snowden. Tuning of mst neurons to spiral motions. *The Journal of Neuroscience*, 14:54–67, 1994.

[13] E.C. Hildreth. *The Measurement Of Visual Motion*. MIT Press, Cambridge, 1983.

[14] E.C. Hildreth, H. Ando, R. Andersen, and S. Treue. Recovering three-dimensional structure from motion with surface reconstruction. *Vision Research*, 35:117–137, 1995.

[15] J.J. Koenderink. Optic flow. *Vision Research*, 26:161–180, 1986.

[16] D.N. Lee. The visual flow field: The foundation of vision. *Philosophical Transactions of the Royal Society of London, B*, 290:169–179, 1980.

[17] D. Marr. *Vision*. W.H. Freeman & Co., San Francisco, 1982.

[18] S.P. Mckee. A local mechanism for differential velocity detection. *Vision Res*, 21:491–500, 1981.

[19] M.C. Morrone, D.C. Burr, and L.M. Vaina. Two stages of visual processing for radial and circular motion. *Nature*, 376:507–609, 1995.

[20] K. Nakayama and J.M. Loomis. Optical velocity patterns, velocity-sensitive neurons, and space perception: a hypothesis. *Perception*, 3:63–80, 1974.

[21] K. Nakayama, G. Silverman, D. McLeod, and J. Mulligan. Sensitivity to shearing and compressive motion in random dots. *Perception*, 14:225–238, 1985.

[22] D.G. Pelli. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10:437–442, 1997.

[23] W.L. Sachtler and Q. Zaidi. Visual processing of motion boundaries. *Vision Research*, 35:807–826, 1995.

[24] G. Sperling, M.S. Landy, B. Dosher, and M. Perkings. Kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception and Performance*, 15:826–840, 1989.

[25] K. Tanaka, Y.Fukada, and H. Saito. Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, 62:642–656, 1989.

[26] A. Treisman and S. Gormican. Feature analysis in early vision: Evidence form search asymmetries. *Psychological Review*, 95:15–48, 1988.

[27] S. Treue and R.A. Andersen. Neural responses to velocity gradients in macaque cortical area MT. *Visual Neuroscience*, 13:797–804, 1996.

[28] J.Y. Wang and E.H. Adelson. Representing moving images with layers. *IEEE Transactions on Image Processing*, 3:625–638, 1994.

[29] W.H. Warren and D.J. Hannon. Direction of self-motion is perceived from optic flow. *Nature*, 336:162–162, 1988.

[30] Y. Weiss and E.H. Adelson. Slow and smooth: a bayesian theory for the combination of local motion signals in human vision. Technical Report AI Memo 1624, Massachusetts Institute of Technology, 1998.

[31] A.L. Yuille and N.M. Grzywacz. A mathematical analysis of the motion coherence theory. *International Journal of Computer Vision*, 3:155–175, 1989.