The perceptual influence of spatiotemporal noise on the reconstruction of shape from dynamic occlusion

Theresa Cooke, Douglas W. Cunningham, and Heinrich H. Bülthoff

Max Planck Institute for Biological Cybernetics, Spemannstrasse 38,
72076 Tübingen, Germany
{firstname.lastname}@tuebingen.mpg.de
http://www.kyb.mpg.de/

Abstract. When an object moves, it covers and uncovers texture in the background. This pattern of change is sufficient to define the object's shape, velocity, relative depth, and degree of transparency, a process called Spatiotemporal Boundary Formation (SBF). We recently proposed a mathematical framework for SBF, where texture transformations are used to recover local edge segments, estimate the figure's velocity and then reconstruct its shape. The model predicts that SBF should be sensitive to spatiotemporal noise, since the spurious transformations will lead to the recovery of incorrect edge orientations. Here we tested this prediction by adding a patch of dynamic noise (either directly over the figure or a fixed distance away from it). Shape recognition performance in humans decreased to chance levels when noise was placed over the figure but was not affected by noise far away. These results confirm the model's prediction and also imply that SBF is a local process.

1 Introduction

The Peacock Flounder can change its coloration such that there are no easily detectable differences in luminance, color, or texture patterns between itself and its surroundings, rendering it almost invisible. When the flounder moves, however, it is immediately and easily visible. This and similar observations suggest that patterns of change over time may be sufficient to visually perceive an object.

This observation was formalized by Gibson [1], who claimed that the pattern of texture appearances and disappearances at the edges of a moving object (i.e., dynamic occlusion) should be sufficient to define that object's shape. Several researchers have shown that this pattern is indeed sufficient for humans to properly perceive not only an object's shape, but also its velocity, relative depth, and degree of transparency [2]. The process of using this dynamic pattern to perceive the shape of an object is referred to as Spatiotemporal Boundary Formation (SBF). The types of transformation that lead to SBF extend well beyond simple texture appearances and disappearances, however, and include changes in the color, orientation, shape, and location of texture elements [3–5]. The use of dynamic information to define a surface avoids many of the problems inherent in static approaches to object perception, and offers a robust way of determining most of the properties of an object from very sparse information while making few assumptions. Machine vision implementations of SBF could be a welcome addition to current object perception techniques.

In a first step towards such an implementation, Shipley and Kellman [5] provided a mathematical proof showing that the orientation of a small section of a moving object (a "local edge segment", or LES) could be recovered from as few as three element transformations. Briefly, each pair of element transformations is encoded as a local motion vector. The vector subtraction of two local motion vectors yields the orientation of the edge. This model predicts that LES recovery, and thus all of SBF, will fail if the elements are spatially collinear, a phenomenon which Shipley and Kellman subsequently psychophysically demonstrated [5]. Likewise, the model suggests that the recovery of the orientation of an LES is sensitive to the spatiotemporal precision of the individual transformations. That is, the more error there is in knowing where and when the elements changed, the more error there will be in the recovered orientation. This implies that SBF should be very sensitive to the presence of dynamic noise (element transformations not caused by dynamic occlusion). Finally, Shipley and Kellman's proof also showed that one should be able to substitute the object's global velocity vector for one of the local motion vectors (which we will refer to as velocity vector substitution), in which case only two element transformations are needed to recover an LES.

Cunningham, Graf and Bülthoff [6–8] revised this proof and embedded it in a complete mathematical framework for SBF. With this framework, the complete global form and velocity of a surface moving at a locally constant velocity can be recovered. The framework consists of three stages. The first stage is similar to Shipley and Kellman's: The orientations of the figure's edges (the LES's) are recovered by integrating element transformations from a local neighborhood in space-time. The elements' locations and the times when they were transformed are encoded relative to each other (i.e., the framework is agnostic on the actual representational format of the changes; it does require them to be encoded as motion vectors). In the second phase, the orientations of the LES's are used in conjunction with the relative spatiotemporal locations of the element transformations to recover the global velocity of the figure. This process, which requires at least two LES's of differing orientations, is mathematically very similar to that used to recover an LES's orientation. If all of the orientations are the same, one can only recover that portion of the global velocity that is perpendicular to the LES's (this is the well-known motion aperture effect). Finally, the global motion of the figure, the orientations of the LES's, and the locations of the element transformations, are used to determine the minimum length of each LES necessary to cause those transformations, as well as the relative locations of the LES's. To complete the process, the LES's may be joined to form a closed contour using a process similar to illusory contour perception (for example, see [9–11]).

Cunningham et al. [12] explicitly tested whether humans can take advantage of the velocity vector substitution process predicted by the model. To do this, they added a set of additional texture elements that had the exact same velocity as the moving shapes. The same set of additional elements was used for all shapes, so they did not provide additional static shape information. Cunningham et al. found that the extra motion information did indeed improve shape identification performance, but only if the new elements were seen as being on the surface of the figure. That is, velocity vector substitution is possible, but only when the extra element motion is seen as belonging to the figure.

So far, all of the model's predictions reflect human performance: Collinearity of the transformations prevents SBF [5], identical orientations of the LES's hinders proper velocity recovery [13], and velocity vector substitution can improve the quality of the recovered shape [12]. What about the prediction that SBF should be strongly affected by dynamic noise (i.e., the presence of transformations that are not caused by dynamic occlusion)? In an inconclusive test of this prediction, Shipley and Kellman [5] performed an experiment that included a condition with a large second element field that jumped around the screen randomly. The presence of this second field impaired SBF (strongly at low element field densities, less so at higher densities). This field may be described as a set of individual elements that flicker on and off, creating appearances and disappearances similar to those produced by dynamic occlusion (in which case the impairment in shape perception demonstrates SBF's sensitivity to dynamic noise). It may also, however, be described as a single element field with a rapidly changing (i.e., Brownian) velocity, and thus the impairment could be accounted for by substituting the Brownian global velocity vector into the LES recovery stage. This latter explanation also accounts for the results in their other experimental conditions, and is consistent with Cunningham et al.'s [12] work on velocity vector substitution, and thus is the most parsimonious explanation.

Both Shipley and Kellman's proof, and Cunningham et al.'s mathematical framework predict, however, that the flickering elements should impair SBF. Here, we explicitly test this prediction by adding a flickering surface texture (i.e., a patch of dynamic noise) to the moving object. Since we can detect the global velocity of dynamic noise fields, and since additional, consistent global velocity information improves SBF, the motion of a flickering surface texture should provide valid global motion information, which should improve SBF. On the other hand, the presence of spurious appearances and disappearances (i.e., flickering elements) near the edges of the object should impair SBF. As a control condition, we examined the effect of a flickering texture that is far away from the moving figure. Since the global motion of a distant texture field does not affect SBF [12], the flickering elements should only affect SBF in the control condition if the spatial integration window for SBF is rather large (i.e., if SBF is more of a "global"than a "local"process).

2 Methods

Ten people were paid 8 Euro per hour to participate in the experiment, which lasted about 30 min. Displays were presented on a 17" CRT monitor. Participants were positioned approximately 50 cm from the screen.

The displays consisted of a 14.6 x 14.6 cm field (visual angle of about 16.3°) of single-pixel, white dots distributed randomly on a black background. One of ten radially monotonic shapes, shown in Figure 1, moved over the random dot field along a circular trajectory of radius 5.72°. The shape completed a single circuit of the trajectory in six seconds. The shapes were identical to those used by Shipley and colleagues in their experiments. This set of shapes has been shown to provide a reliable means of determining which variables affect SBF [2].

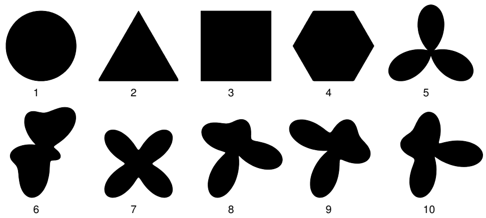

Fig. 1. The ten shapes used in the experiment.

This experiment only used the “unidirectional” type of displays: Whenever the leading edge of the form moved over a white dot, the dot was transformed to black and "disappeared" from the display. When the trailing edge of the form reached the dot, the dot was changed back to white, thus "reappearing" in the display. A second type of display that is typically used, called a “bidirectional” display, is identical to a unidirectional display, with the sole exception that elements may either appear or disappear along any edge (i.e., half of the elements are only visible outside of the figure, as in the unidirectional displays, and half are only visible inside the figure). Well-defined shapes are seen in the bidirectional displays, despite the absence of any form of shape information except dynamic occlusion [5]. Bidirectional displays were not used in the current experiment for theoretical reasons (i.e., there are some concerns about surface formation, the direction of surface binding, and the role of velocity vector substitution in bidirectional displays, see Cunningham et al. [12] for more on this topic). The number of dots was systematically varied: The background had 100, 200, or 400 elements.

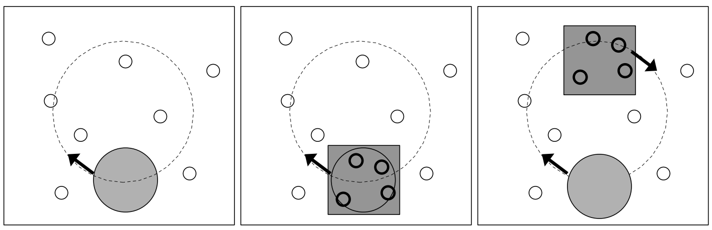

Fig. 2. Sketch of the three experimental conditions: a) "noise-free": the occluder moves through the random dot field along a circular path; b) "noise near": the noise pattern, represented by dark black dots inside a square, is superimposed on the moving figure; c) "noise far": the moving figure and noise pattern are separated by 180°.

Three noise conditions were used: a condition without noise, a condition where dynamic (i.e., spatiotemporal) noise was placed near the object, and a condition where dynamic noise was placed far from the object (see Figure 2). The dynamic noise pattern was a set of white dots which appeared at random locations inside a virtual box of fixed size. The size of the box was chosen such that it circumscribed the largest object. In this way, the object could not be identified using the size or the shape of the noise pattern. The number of dots placed inside the box was 15% of the number of dots in the background. The noise dots had a limited lifetime – the location of the noise dots was changed every four frames. Note that the noise dots themselves never moved. The new location of each dot was chosen to be within the virtual box. In the “noise-near” condition, the box containing the noise dots was superimposed on the moving figure. In the “noise-far” condition, the noise pattern was placed 180°away from the figure along the circular trajectory.

The experiment consisted of a single block of 90 randomized trials (10 figures x 3 densities x 3 noise conditions). Participants were asked to identify the shape moving in each display using a ten-alternative forced-choice task. Static images of the ten possible choices were shown at all times on the left side of the monitor. Each shape moved around its circular trajectory until the participant responded or until the shape completed two cycles through the trajectory, whichever came first. If the participant did not respond before the end of the second cycle, the display was cleared (except for the 10 reference shapes on the side of the screen), and remained blank until the participant responded.

3 Results

3.1 Effect of Density

In both the "noise-free" and "noise far" conditions, there was a significant effect of density, consistent with previous work on SBF. All tests were performed using a two-tailed t-test for independent samples with equal variances; all conditions for using these tests were met. Average performance was significantly higher at density 200 than at density 100 (all $t's(18)>4.1$, all $p's<0.001$), and significantly higher at density 400 than at density 200 (all $t's(18)>2.1$, all $p's<0.05$). There was no effect of density for the noise-near condition – performance was at chance (the performance level expected from blind guessing) at all density levels. This is almost certainly due to a floor effect, meaning that improvements in performance would probably have been observed had the task been easier or the number of choices greater.

3.2 Effect of Noise

At no density level was mean performance in the "noise near" condition significantly different from chance (all $t's(9)<=1.8$, all $p's>0.1$). Moreover, mean performance in the "noise near" condition differed significantly from mean performance in the "noise-free" condition at all density levels (all $t's(18)>=2.60$, all $p's<0.05$). At the two higher density levels, mean performance in the "noise near" condition also differed significantly from mean performance in the "noise far" condition ($t(18)=7.19$, $p<0.001$ at density 200; $t(18)=12.38$, $p<0.001$ at density 400). At a density of 100, the difference was not significant ($t(18)=1.80$, $p>0.05$ (n.s.)), but performance in the "noise far" condition did vary significantly from chance at this density level ($t(9)=2.75$, $p>0.01$).

The mean accuracies in the "noise-free" and "noise far" conditions were not significantly different at any density level (all $t's(18)<=0.7$, all $p's>0.2$).

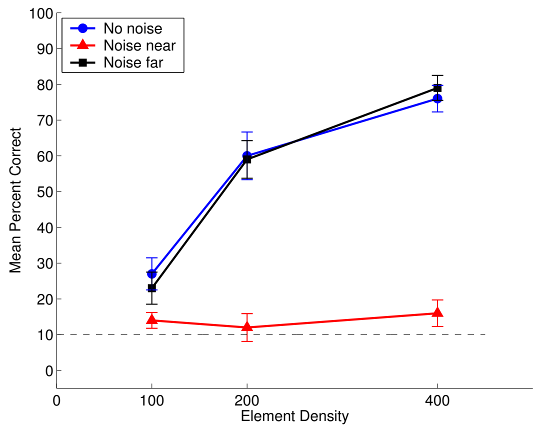

Fig. 3. Shape identification accuracy plotted as a function of element density for the three conditions. Error bars represent the standard error. The dotted line represents chance performance.

4 Conclusions and Discussion

At all density levels, shape identification performance was reduced to chance when spatiotemporal noise was placed near the figure, whereas it was unaffected by noise placed far away. The fact that the dynamic noise in the "noise-near" condition prevented accurate shape recognition suggests that the spurious appearances and disappearances were being treated as dynamic occlusion signals. This would, according to the model, impair LES recovery and prevent shape perception. Since shape recognition performance was at chance level in the present experiment, it is possible that the presence of dynamic noise prevented SBF from occurring at all. That is, the low signal-to-noise ratio may be a signal that the entire SBF process should not be performed. In Shipley and Kellman's [5] experiment with dynamic noise, however, the noise merely reduced recognition accuracy. Since the same shapes, task, density levels, and shape velocity were used in both experiments, the differences in shape recognition performance are probably due to the differences in the dynamic noise. The dynamic noise patch was much denser in the present experiment, and was focused around the shape itself. Thus, it seems that the individual noise signals are being integrated with the dynamic occlusion transformations, which produces LES's that are incompatible with the true shape of the moving figure, which in turn leads to failures in the subsequent global form reconstruction. This suggests that one might use dynamic noise to carefully probe the exact characteristics of LES recovery. For example, one might vary the location, density, or distribution of noise to precisely determine the spatial and/or temporal integration windows, element grouping processes, or global form reconstruction mechanism of SBF.

It should be possible to implement an iterative consistency filter to remove at least some of the inconsistent LES's, reducing the sensitivity of SBF to dynamic noise. Although the human visual system does not seem to employ such a filter, machine vision implementations of SBF (such as that by Cunningham et al. [7, 8]) might benefit from such a filter.

The insensitivity of SBF to the velocity of the dynamic noise patch in the noise-far condition confirms Cunningham et al.'s [12] claim that only global motion seen as coming from a surface's texture affects SBF. The insensitivity of SBF to the spurious transformations produced by distant dynamic noise patch provides evidence that SBF is a strictly local process. This finding places convenient restrictions on the LES's recovery stage, and eases the computational overhead that would be involved in an iterative filter to remove inconsistent LES's.

Perceptually, it was clear in these displays that the dynamic noise patch was moving coherently *as a whole*, yet this global motion information did not seem to help SBF. It is possible that the global motion pattern did help, but that this positive contribution was outweighed by the detrimental effect of the spurious flickering of the noise patch. Another interesting possibility is that the improvement in SBF produced by adding a coherent surface texture found by Cunningham et al. [12] was not due to the motion of the surface *texture*, but to the motion of the surface *texture elements*. Since the individual elements in the noise patches did not move, there was no motion to disturb SBF (static element fields imposed on dynamically defined figures do not affect SBF very strongly [14]).

The results presented here confirm some previously untested predictions of Cunningham et al.'s model of SBF and provide additional constraints on potential computational implementations of SBF. It seems that SBF is a robust method for extracting most properties of a moving object from very sparse information while making few assumptions about the structure of the world.

References

1. J. J. Gibson, G. A. Kaplan, H. N. Reynolds, Jr., and K. Wheeler. The change from visible to invisible: A study of optical transitions. *Perception & Psychophysics*, 5:113 – 116, 1969.
2. T. F. Shipley and D. W. Cunningham. Perception of occluding and occluded objects over time: Spatiotemporal segmentation and unit formation. In T. F. Shipley and P. J. Kellman, editors, *From fragments to objects: Segmentation and grouping in vision*, pages 557 – 585. Elsevier Science, Oxford, 2001.
3. T. F. Shipley and P. J. Kellman. Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion? *Spatial Vision: Special Issue: In honor of Bela Julesz*, 7:323 – 339, 1993.
4. T. F. Shipley and P. J. Kellman. Spatiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General*, 123:3 – 20, 1994.
5. T. F. Shipley and P. J. Kellman. Spatiotemporal boundary formation: The role of local motion signals in boundary perception. *Vision Research*, 37:1281 – 1293, 1997.
6. D.W. Cunningham, A.B.A. Graf, and H.H. Bülthoff. A relative encoding approach to modelling spatiotemporal boundary formation. *Journal of Vision, VSS Abstract*, page 120, 2002.
7. D.W. Cunningham, A.B.A. Graf, T. Cooke, C. Wallraven, and H.H. Bülthoff. The perception of objects from dynamic occlusion: A relative encoding implementation. under review, 2004.
8. T. Cooke, D. W. Cunningham, and C. Wallraven:. Local processing in spatiotemporal boundary formation. *Proceedings of the 7th Tübingen Perception Conference*, page 65, 2004.

9. S. Grossberg and E. Mingolla. Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92:173–211, 1985.
10. T. F. Shipley and P. J. Kellman. Boundary completion in illusory contours: Interpolation or extrapolation? *Perception*, 32:985–999, 2003.
11. Lance R. Williams and David W. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, 9(4):837–858, 1997.
12. D. W. Cunningham, T. F. Shipley, and P. J. Kellman. Interactions between spatial and spatiotemporal information in spatiotemporal boundary formation. *Perception & Psychophysics*, 60:839 – 851, 1998.
13. H. Wallach. Über visuell warhgenomme bewegungsrichtung. *Psychologische Forshcung*, 20:325 – 380, 1935.
14. D. W. Cunningham, T. F. Shipley, and P. J. Kellman. The dynamic specification of surfaces and boundaries. *Perception*, 27:403 – 416, 1998.