Vision Research 126 (2016) 131–142

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Modeling spatiotemporal boundary formation

Gennady Erlikhman [*], Philip J. Kellman

Department of Psychology, University of California, Los Angeles, United States

ARTICLE INFO

Article history:
Received 9 February 2015
Received in revised form 23 March 2015
Available online 12 April 2015

Keywords:
Boundaries
Illusory contours
Perceptual organization
Grouping
Perceptual model

ABSTRACT

Spatiotemporal boundary formation (SBF) refers to perception of continuous contours, shape, and global motion from sequential transformations of widely separated surface elements. How such minimal information in SBF can produce whole forms and the nature of the computational processes involved remain mysterious. Formally, it has been shown that orientations and motion directions of local edge fragments can be recovered from small sets of element changes (Shipley & Kellman, (1997). *Vision Research*, *37*, 1281–1293). Little experimental work has examined SBF in simple situations, however, and no model has been able to predict human SBF performance. We measured orientation discrimination thresholds in simple SBF displays for thin, oriented bars as a function of element density, number of element transformations, and frame duration. Thresholds decreased with increasing density and number of transformations, and increased with frame duration. An ideal observer model implemented to give trial-by-trial responses in the same orientation discrimination task exceeded human performance. In a second group of experiments, we measured human precision in detecting inputs to the model (spatial, temporal, and angular inter-element separation). A model that modified the ideal observer by added encoding imprecision for these parameters, directly obtained from Exp. 2, and that included two integration constraints obtained from previous research, closely fit human SBF data with no additional free parameters. These results provide the first empirical support for an early stage in shape formation in SBF based on the recovery of local edge fragments from spatiotemporally sparse element transformation events.

1. Introduction

A primary goal of the visual system is to use information in reflected light to perceive objects and surfaces. Crucial among the processes involved is detection of edges and surface boundaries, for which there are many cues, including discontinuities in luminance contrast, color, stereoscopic disparity, and texture. However, these cues may sometimes be insufficient, when depth differences are below threshold, in poorly lit environments, or where only sparse surface elements are visible. In such cases, surface boundaries can often be revealed by object or observer motion. Dynamic cues, especially the accretion and deletion of texture (Gibson et al., 1969), can provide sufficient information for the segmentation of similarly or sparsely textured surfaces and can result in the perception of boundaries, surfaces, and global motion (Kaplan, 1969; Andersen & Cortese, 1989; Gibson et al., 1969; Yonas, Craton, & Thompson, 1987; Stappers, 1989; Shipley & Kellman, 1993, 1994).

Although accretion and deletion of texture has been described primarily as a cue to relative depth (Gibson et al., 1969), it has also been noted that it produces perception of shape in the absence of any other cues to shape (Andersen & Cortese, 1989; Gibson et al., 1969; Shipley & Kellman, 1993, 1994). These latter phenomena pose a mystery. The perception of continuous illusory contours (and the shapes they delineate) across empty surface regions between elements does not obviously follow from the perception of occlusion of an element.

Shipley and Kellman (1993, 1994) found that gradual occlusion of elements was not even necessary, as discrete element disappearance also produces perceptions of boundaries and surfaces across gaps. Further, no form of accretion and deletion of texture elements, continuous or discrete, is needed. The visual system appears to use *any* abrupt change in local elements as inputs to a process that produces perceived edges, form, and global motion. Changes in element orientation, shape, color, or position all produced these effects, and they labeled this more general process of perception of continuous illusory boundaries and global form from sequential changes in local surface elements *spatiotemporal boundary formation* (SBF).

[*] Corresponding author.
E-mail address: gerlikhman@unr.edu (G. Erlikhman).

http://dx.doi.org/10.1016/j.visres.2015.03.016

How do local element changes produce the continuous boundaries seen in SBF? It has been proposed that shape in SBF depends on two processing stages (Shipley & Kellman, 1994, 1997). First, information from sets of element changes in small neighborhoods somehow produce local, oriented edge fragments. Second, these edge fragments connect to each other across gaps according to well-known interpolation processes that operate in the perception of illusory and occluded contours (Fantoni & Gerbino, 2003; Grossberg & Mingolla, 1985; Kanizsa, 1979; Michotte, Thines, & Crabbe, 1964; Kellman & Shipley, 1991; Palmer et al., 2006). Whereas the second stage involves processes that are well-understood, the first stage has remained mysterious. Shipley and Kellman (1994) showed mathematically that a local orientation could be derived from three sequential non-collinear element transformations. Little empirical research, however, has examined SBF with single edges and relatively few elements. Virtually all previous studies of SBF have used closed objects with smooth contours as stimuli (although see Barraza & Chen, 2006). Recently, we demonstrated that individual, oriented, illusory edge fragments can be recovered from sparse displays (Kellman et al., 2012). These results support the two-level theory of SBF, specifically in implicating a process that recovers local oriented edge fragments. These fragments are likely the basic units from which larger shapes are constructed in SBF.

Here we sought to develop and test a process model of how such edges are extracted. We implemented and tested an ideal observer model of edge extraction in SBF displays, based on the idea that triplets of sequential element transformations can provide an estimate of a local, oriented edge fragment. In Experiment 1, we measured orientation discrimination thresholds for SBF-defined edges across a variety of display properties. Human performance was much worse than the ideal observer model. Unlike the model, human performance may involve noise in registering relevant inputs as well as limits on information accumulation. In a second experiment, we used separate paradigms to measure noise in human registration of basic input features, such as inter-element separation. A model that incorporated simple information accumulation constraints and the measured spatial and temporal noise parameters in Experiment 2 was able to accurately predict human performance from Experiment 1 across all tested display conditions.

1.1. Background: SBF displays and models

In this section, we briefly review SBF phenomena and prior models. Fig. 1 shows an example of an SBF display. The dotted line defines the boundary of a virtual object. The elements are always stationary and the virtual object moves across the display. As the object moves, elements that fall within the boundary change in some property, such as color. The change is discrete, and the percept is of a moving figure with clear boundaries. In *unidirectional transformations*, elements initially have one value (e.g., white dots on a black background) and when they become encompassed within the virtual region, they change to a different value (e.g., white dots turn blue). Upon exiting the region, elements revert to their original value (e.g., blue dots revert to white). In *bidirectional transformations*, elements are randomly assigned one of two values and switch to the other upon entering or exiting the boundary of the moving object. For example, with blue and white dots on a black background, blue dots turn white upon entering the virtual region, and white dots turn blue. SBF occurs across a wide variety of parameters, with the precision of shape perception depending on element density, luminance differences between elements, the velocity of the virtual region, and frame duration (Andersen & Cortese, 1989; Cicerone et al., 1995; Shipley & Kellman, 1994).

In SBF, no single frame has visible edges of a shape. In some unidirectional transformation displays, there will be a region of elements having a different feature value from surrounding elements, but the shape of this region is not well specified. Other unidirectional transformations, such as local element motion, as well as all bidirectional transformations, offer no information in any static frame about shape or about any affected region. Because elements transform all at once, there is no oriented contour information as might be given by gradual occlusion of an object or texture element. Thus, in SBF, local edges are not given by any of the standard cues for edge perception. Moreover, even for a mechanism attempting to extract local edge fragments from local changes in element properties, SBF displays pose a difficult variant of the aperture problem (Adelson & Movshon, 1982; Wallach, 1935), what has been referred to as the "point aperture problem", in which *neither* the orientation nor velocity of an edge are directly given in the stimulus (Prophet, Hoffman, & Cicerone, 2001; Shipley & Kellman, 1994, 1997). In the point aperture problem, there are no oriented edge fragments given in the stimulus. The visual system must simultaneously recover *both* the orientation and motion of a local edge from sparse and discrete element transformations.

A solution to the point aperture problem was proposed by Shipley and Kellman (1994, 1997). Given the positions and times of occurrence of three, non-collinear element transformations, the orientation of an edge that caused those transformations can be computed assuming a constant edge velocity and orientation (Shipley & Kellman, 1997). An intuition for the proof appears in Fig. 2. Fig. 2a depicts a sequence of element transformations (labeled 1, 2, and 3) caused by a moving edge. When two elements transform (in this case, disappear and reappear) in succession, a transformation vector, $\mathbf{v}_{12}$, is formed between them. The magnitude of the vector is determined by the spatial and temporal separation of the transformations. We use the term "transformation

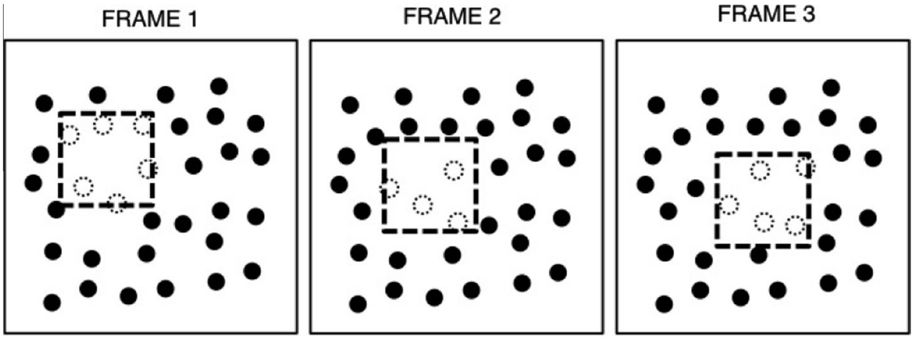

Fig. 1. Depiction of a square "virtual region" moving over a field of circular black elements. All elements inside the square region are in one state (white) and all those outside are in another (black). As the square moves (frames 2 and 3), elements entering and exiting the region change states. The resulting percept is of a moving region with crisply defined illusory contours.

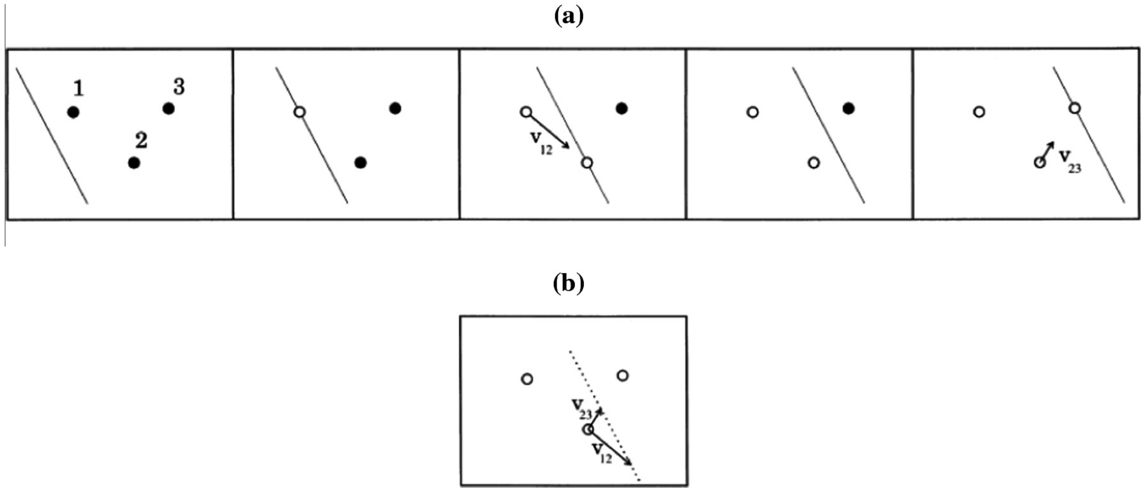

Fig. 2. A sequence of frames in which a moving edge successively transforms three elements (changing from black to white). (a). Three elements disappear, one at a time. $\mathbf{v}_{12}$ and $\mathbf{v}_{23}$ are transformation vectors defined by the spatial and temporal separation between elements. (b). Transformation vectors $\mathbf{v}_{12}$ and $\mathbf{v}_{23}$ can be combined to define the orientation of the moving edge. Figure from Shipley and Kellman (1997). Spatio-temporal Boundary Formation: the Role of Local Motion Signals in Boundary Perception. *Vision Research*, *37* (10), 1281–1293.

vector” instead of motion vector to emphasize that apparent motion is *not* seen between individual elements during SBF. The transformation of a third element defines a second transformation vector, $\mathbf{v}_{23}$, between the second and third elements. If the tails of these two transformation vectors are placed on the same point (Fig. 2b), then the orientation of the vector connecting their heads ($\mathbf{v}_{12}$–$\mathbf{v}_{23}$) has the orientation of the illusory edge, provided that the edge was moving at a constant velocity and had a constant orientation between transformation events.

The orientation of the illusory edge, θ, is given by the following equation:

$$\theta = \tan^{-1}\left(\frac{\mathrm{v}_{23} * \sin\varphi_{23} - \mathrm{v}_{12} * \sin\varphi_{12}}{\mathrm{v}_{23} * \cos\varphi_{23} - \mathrm{v}_{12} * \cos\varphi_{12}}\right) \qquad (1)$$

where φ_{ij} is the angle formed between a horizontal line passing through element i and a line connecting element i to element j, and v_{ij} is the magnitude of the transformation vector between the two elements. v_{ij} can be computed from the distance between two elements (A_{ij}) and the time between the transformations of those elements (ΔT_{ij}):

$$v_{ij} = \frac{A_{ij}}{\Delta T_{ij}} \qquad (2)$$

In the classical aperture problem (Adelson & Movshon, 1982; Nakayama, & Silverman, 1988a, 1998b; Shimojo, Silverman, & Nakayama, 1989), the orientations of the edges or gratings are given by contrast information. In SBF, the edge segment *itself* is not an input to the computation; it is what is being recovered. Once edge orientation is recovered, the motion direction of the edge is itself ambiguous (Shipley & Kellman, 1994). Motion direction of the recovered edge segment can be solved if several differently oriented segments along the object boundary are recovered, as in the classical aperture problem.

The model of edge orientation recovery makes several assumptions. First, at least three transformation events are needed. Orientation is ambiguous for any two events because an infinite number of combinations of edge orientations and velocities could produce those two transformations. Second, the three elements cannot be collinear. Third, orientation and velocity of the moving edge must be constant between transformation events.

For many of the element transformation types that produce SBF, any pair of sequential, nearby element changes, if viewed in isolation, would produce perception of apparent motion. A remarkable property of SBF is that these local, nearest neighbor apparent motions (c.f. Ullman, 1979) are *not* what is seen in SBF. Instead larger moving boundaries and shapes are seen (from appropriate collections of element changes). Thus, it appears that motion-like signals function as inputs to SBF, but are used in a different way from what they would signal in isolation. Several findings support the hypothesis that motion-like signals, or vectors relating pairs of element transformations serve as the input to an edge extraction process. For example, SBF can be disrupted by the addition of spurious flickering or moving background elements (Cooke, Cunningham, & Bülthoff, 2004; Cunningham, Shipley, & Kellman, 1998; Shipley & Kellman, 1997). Contour clarity in SBF also depends on the relative contrast of elements. When element transformations are isoluminant color changes, illusory contour perception is greatly reduced (Cicerone et al., 1995; Miyahara & Cicerone, 1997). First-order motion perception is also poor under isoluminance (Cropper, 2005; Cropper & Derrington, 1994; Derrington & Henning, 1993), suggesting that the similar motion mechanisms may be affected in SBF as well.

Despite behavioral evidence in support of a basic edge-extraction process that uses motion-like transformation vectors, no working model has previously been implemented or tested that takes an SBF display as input and produces a local edge orientation as output. Part of the difficulty has been that most SBF displays have used 2D shapes as virtual objects, the recovery of which would require not only this first edge extraction stage, but also a second stage in which those edges are integrated and missing regions are interpolated. We created a display in which the SBF-defined shape was a single, thin, oriented bar that translated horizontally across the screen. In Experiment 1, subjects performed an orientation discrimination task with the bar. Because the bar could be treated as a single edge, the model could be directly applied to the displays to compute the bar's orientation. The model's orientation discrimination threshold could then be computed from these estimates and directly compared to human performance.

2. Experiment 1

A virtual bar moved across a field of black, circular elements (Fig. 3). Whenever the bar passed the midpoint of an element, that element disappeared (became white) all at once and remained invisible (white) for two frames, at which point it reappeared. Human orientation discrimination thresholds were measured as

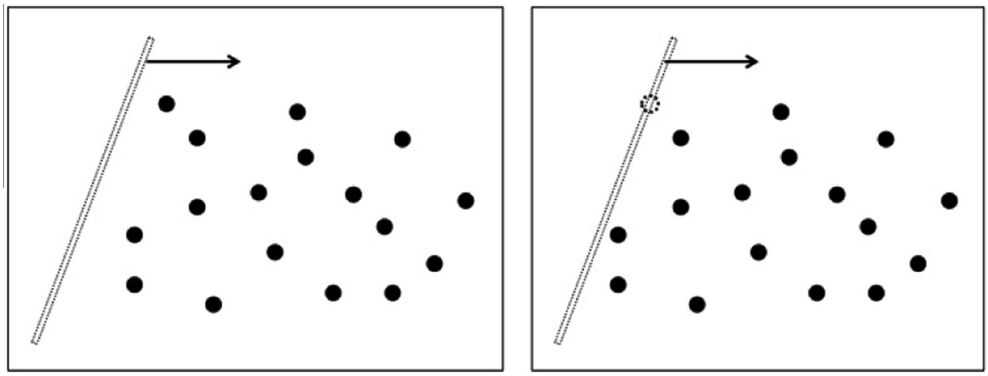

Fig. 3. Illustration of stimuli used in Experiment 1. An invisible, oriented bar moved laterally across a field of black elements on a white background. Whenever it passed the midpoint of an element, that element disappeared (became white; indicated by dashed circle in second panel) all at once. The element remained white for two frames and then reappeared (became black). The perception was of a moving, illusory, white bar.

a function of several display parameters: element density, number of element transformations, and frame duration. These manipulations have previously been shown to affect the perception of illusory contours in SBF (Shipley & Kellman, 1994). If the model is correct, it should accurately predict performance under a variety of display settings and be affected by the same properties that affect human performance. The model described in Eq. (1) was used to predict edge orientation on a trial-by-trial basis in simulated experimental trials. The orientation estimates were then submitted to the same staircase procedure as human data to determine the bar's orientation on subsequent trials and to estimate an orientation discrimination threshold. The model and human orientation discrimination thresholds were then compared as a function of the manipulated display properties. We discuss the properties of the model after presenting the behavioral results.

2.1. Method

2.1.1. Participants

Subjects were 45 students from the University of California, Los Angeles, split into groups of 15 for each of the three experimental conditions. Subjects were compensated with course credit for participating. All reported having normal or corrected-to-normal vision. The subjects were naïve to the purposes of the experiment. All subjects provided informed consent and this work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.1.2. Design

A between-subjects design was used to test the effects of three display properties on orientation discrimination of SBF-defined edges. Displays varied in element density (number of elements per square region), number of transformation events, or frame duration. Each group of subjects was exposed to only one display manipulations. All subjects judged whether an SBF-defined edge was tilted clockwise or counterclockwise away from vertical. Orientation sensitivity was measured for six element densities, six element quantities, and three frame durations.

2.1.3. Apparatus

Stimuli were created and displayed using the MATLAB programming language and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Stimuli were presented on a Viewsonic G250 CRT monitor, which was powered by a MacPro 4 with a 2.66 GHz Quad-Core Intel Xeon processor and an NVidia GeForce GT120 graphics card. The monitor was set to a resolution of 1024 × 768 pixels and a refresh rate of 60 Hz.

2.1.4. Displays

Displays contained black, circular elements with a diameter of 10 pixels (0.25 degrees of visual angle) on a white background. The elements were placed within a 614.4 pixel by 614.4 pixel region (15.19° × 15.19°) centered on the computer monitor. The elements were pseudo-randomly arranged by dividing the display area into a grid of equally sized regions and placing a single element at a random position within each region. This placement method ensured that there were no large areas in the display that lacked elements and also prevented their overlap while preserving a somewhat uniform distribution over the entire display (cf. Shipley & Kellman, 1993, 1994).

A one-pixel-wide bar was specified that spanned the height of the display. On each frame, the bar moved laterally 5 pixels (0.125 deg/frame, 7.5 deg/s). Whenever the bar passed the midpoint of an element, that element disappeared (became white) for two frames (33.2 ms) and the bar paused. After two frames, the element reappeared (became black) and the bar continued moving. Elements appeared and disappeared discretely without gradual occlusion. The resulting percept was of a horizontally translating, illusory bar. Whether the bar started on the left or right side of the display was randomized across trials. The trial lasted until the bar reached the opposite end of the screen, so each element transformed only one time. A new arrangement of elements was generated for every trial.

On each trial, the bar was tilted clockwise or counterclockwise with respect to the vertical. The degree of tilt was set by an adaptive staircase procedure (Psi method (Kontsevish & Tyler, 1999) implemented in the Palamades Toolbox (Prins & Kingdom, 2009)) that was used to find the 75% orientation discrimination threshold. Whether the bar was rotated clockwise or counterclockwise was randomized across trials.

We independently manipulated: element density, number of transformation events, and frame duration. Element density was varied by drawing 9, 16, 25, 36, 49, or 64 elements in the display area, corresponding to densities of 0.04, 0.07, 0.11, 0.16, 0.21, and 0.28 elements per squared degree visual angle. A separate staircase was used for each density. The six staircases were interleaved and terminated after 50 trials.

In the event number condition, element density was held constant at 0.28 elements per squared degree of visual angle (the highest density in the density condition). Each display contained 64 elements. The trial lasted until the illusory bar came into contact with 9, 16, 25, 36, 49, or 64 elements. Starting horizontal position and motion direction of the bar were randomized with the constraint that there would be enough elements in the direction of motion that would allow for the required number of element contacts. As with density, six interleaved staircases were used.

In the temporal condition, 64 elements were placed with the highest density used from the density and event conditions. Frame durations were 16.7, 33.3, or 66.7 ms. Subjects were allowed to respond at any point during the trial. Three interleaved staircases were used, one for each frame duration. The shortest frame duration was the same frame duration that was used in the other conditions. As such, there was one display type that was identical in across all three conditions (64 elements, 64 events, 16.7 ms frame duration).

2.1.5. Procedure

Subjects sat in a dark room at a distance of 89.5 cm from the monitor. The only illumination came from the monitor. Subjects were instructed that they would be making orientation judgments about slanted edges and were shown examples of real edges that were tilted clockwise and counterclockwise, at which point they began the experiment. Before beginning experimental trials, subjects first performed 10 practice trials at the highest element density and quantity. After each stimulus presentation, a response screen appeared asking whether the line was tilted clockwise or counterclockwise. Subjects made a response by pressing a key on the keyboard. Feedback was provided after each practice trial. Once complete, subjects were told that they would receive no further feedback. Rest breaks were provided every 100 trials.

2.2. Results and discussion

Orientation discrimination thresholds are shown in Fig. 4 (black lines). The 75% correct orientation discrimination thresholds were computed for each subject for each condition and averaged across subjects. For each condition, data were submitted to a within-subject, one-way ANOVA to test for the effect of the manipulated display property. Increasing density decreased thresholds with the highest threshold of 19.08° for the lowest density and 3.17° for the highest density (Mauchly's test: $\chi^2(14) = 43.85$ $p < 0.001$, Greenhouse-Geisser $\varepsilon = 0.37$, $F(1.86,\ 26.06) = 87.77$, $MSE = 16.65$, $p < 0.001$, $\eta^2_p = 0.86$). Similarly, increasing the number of element transformations decreased thresholds (Mauchly's test: $\chi^2(14) = 71.25$ $p < 0.001$, Greenhouse-Geisser $\varepsilon = 0.33$, $F(1.64,\ 22.91) = 8.35$, $MSE = 19.6$, $p = 0.003$, $\eta^2_p = 0.37$). Increasing the inter-frame interval increased thresholds ($F(2,\ 28) = 7.78$, $MSE = 2.37$, $p = 0.002$, $\eta^2_p = 0.36$).

In the element transformation event quantity condition, displays with only nine element transformations had the highest thresholds of 7.95°. Displays with 16 or 25 element transformations had slightly lower thresholds of 4.42° and 3.76° respectively. Finally, displays with 36 or more element transformations had similar thresholds: 3.25°, 2.97°, and 3.12° for 36, 49, and 64 transformations respectively. This pattern suggests that orientation discrimination performance reached an asymptote at around 16 or 25 transformations and did not improve with additional transformations. To confirm this leveling off of performance, thresholds were compared when looking only at 16–64 events and when looking only at 25–64 events. When the 9-event condition was excluded, there was still a marginal effect of the number of events (Mauchly's test: $\chi^2(9) = 28.73$ $p < 0.001$, Greenhouse-Geisser $\varepsilon = 0.50$, $F(2.00,27.97) = 8.35$, $MSE = 3.91$, $p = 0.085$, $\eta^2_p = 0.16$). However, when both the 9- and 16-event conditions were excluded, there was no significant difference in thresholds for the remaining event quantities (Mauchly's test: $\chi^2(5) = 15.80$ $p = 0.008$, Greenhouse-Geisser $\varepsilon = 0.57$, $F(1.70,23.86) = 1.55$, $MSE = 2.00$, $p = 0.233$, $\eta^2_p = 0.10$). Orientation discrimination performance therefore appeared to level off at around 25 events.

For frame duration, average thresholds were similar for the two fastest durations (3.90° and 3.66° respectively) and worse for the longer, 66.7 ms frame duration (5.70°). Displays containing 64 events, with the highest density and shortest frame rate appeared in all three conditions. Thresholds were not significantly different for these displays across the three conditions ($ps > 0.05$).

The results concur with previous findings that element density and frame duration affect shape perception in SBF (Shipley & Kellman, 1994). The present findings go further in showing that single edges can be recovered in SBF even when there is no complete 2D shape with varying orientations. An early formal analysis (Shipley & Kellman, 1994) showed that for a 2D shape, information along differently oriented parts of the boundary, along with a constant velocity constraint, could allow recovery of orientation in SBF. The present findings indicate that varying orientations along a virtual object are not required for SBF to operate. Another interesting finding was that in the event quantity condition, performance continued to increase as a function of the number of events up to approximately 25 events, after which performance leveled off. This result suggests an important constraint on information accumulation in human SBF performance.

2.3. Ideal observer model

The ideal observer model described by Eq. (1) was used to predict bar orientation on a trial-by-trial basis for each of the conditions in Experiment 1. On each trial, the relative distances (A_{ij}), angular relationships (φ_{ij}), and timing (ΔT_{ij}) of element transformations were recorded for all elements directly from the displays

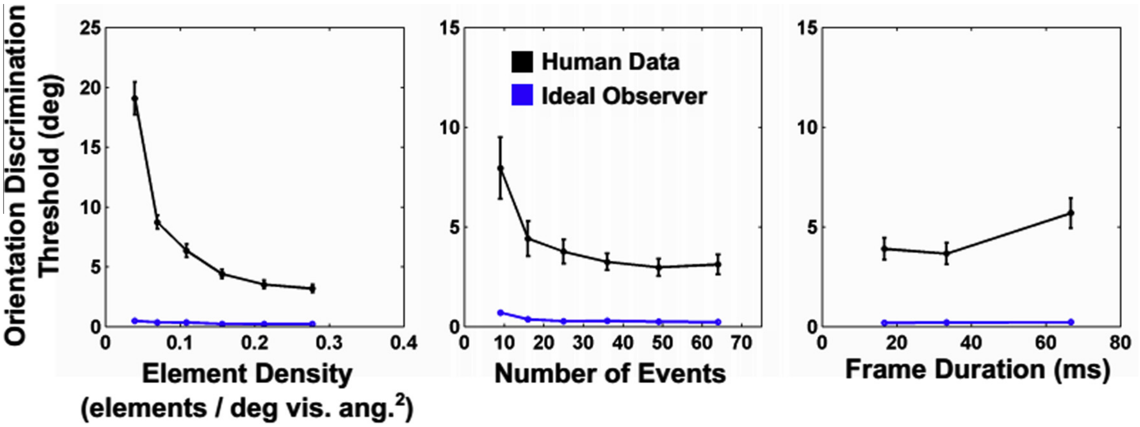

Fig. 4. Average orientation discrimination thresholds for three display conditions tested in Experiment 1. Thresholds are shown as a function of element density, number of events, and frame duration in the graphs going from left to right. Human performance data are shown in black; ideal observer performance is in blue. Bars indicate standard error of the mean. Note the difference in scale of the y-axis between the first vs. second and third graphs.

by simulating the motion of the bar. This resulted in a sequence of events that corresponded to the bar sequentially interacting with each element. The sequence of events was divided into groups of three and edge orientation was computed for each such triplet. From n elements in a display, $n-2$ triplets were created. Each triplet yielded an estimate of bar orientation, θ. All elements except for the first and last appeared in multiple triplets. The median of the orientation estimates in a single trial was used to generate a "clockwise" or "counterclockwise" response. If the median was 90°, one of the two responses was chosen randomly. The responses were then submitted to the same staircase procedure used by human subjects. This process was repeated for each experimental condition and display setting. Importantly, while the model output was an orientation for a single triplet of elements, the comparison to human data was at the level of discrimination thresholds.

The model was able to predict edge orientation very accurately, producing thresholds below one degree for all densities (blue lines, Fig. 4). However, in examining individual orientation estimates derived from a triplet, there was some deviation from true orientation. Average orientation estimate error was 1.85° per triplet. When the median was taken across all orientation estimates computed from all triplets in a single trial, error was 0.37°, 0.28°, 0.24°, 0.20°, 0.17°, and 0.14° for the six element densities from smallest to largest respectively. Error was reduced for higher density displays because there were more triplets that contributed to the final estimate. The model's performance may have been imperfect because the bar advanced in discrete steps of 5 pixels every 16.7 ms (i.e., every frame). This introduced error in the amount of time between element transformation events, which could only be in multiples of the frame rate. To test this explanation, a separate set of simulations was run for which the velocity of the bar was used to compute the time when an element should have transformed. Using these "true" times, average orientation estimate error was less than 0.1° per triplet. The model is therefore able, in principle, to perfectly determine edge orientation from three element transformation events. In all subsequent modeling, the timing correction was not applied because true bar velocity cannot be known *a priori*. Even without the correction, however, the model's median orientation estimates differed little (less than 0.5°) from true orientation, and the final model thresholds were well below those of human observers, even for the highest density.

It is possible that simultaneous element transformation events could have affected both human and model performance. Simultaneous events could be used to perform the task perfectly: because the stimulus was a thin bar, simultaneous events could only have occurred if the orientation of the edge was the same as the angle between the event positions. Knowing the angle would therefore be sufficient to determine whether the edge was oriented clockwise or counterclockwise. This strategy would be particular to the displays used in this experiment: if the contour of the virtual object was curved or composed of more than a single edge, a straight line connecting the positions of simultaneous events on different parts of the curve or on different edges would not correspond to the shape's contour. Nevertheless, if observers discovered that they could use simultaneous events to do the task, then one might have expected performance to have been near-perfect, especially for higher densities which had the highest frequencies of simultaneous events on a per-trial basis. On average, there were 0.21, 0.74, 1.93, 4.06, 7.56, and 12.73 simultaneous events per trial for each of the six element densities respectively. However, even at the highest density, average human thresholds were around 3.5°. If two simultaneous events over the course of a trial were sufficient to perform the task, we would have expected better performance. Furthermore, because the virtual edge spanned the height of the display, simultaneous events were often far apart. The average distance between simultaneous events for the highest density was 3.38°. Given that simultaneous events lasted for only 33.2 ms and that the elements were small, it would have been difficult for observers to detect them at all.

As a further check, we also performed a control experiment (not reported here) in which only a single element disappeared on every frame. Subjects reported seeing an illusory edge and orientation discrimination thresholds were very similar to those found in Experiment 1. The model was also able to perfectly predict edge orientation in these displays. Human and model performance therefore did not depend on the presence of simultaneous transformation events.

3. Experiment 2

Experiment 1 indicated that observers very accurately discriminate between edge orientations of illusory edges defined solely by SBF and that their sensitivity depended on the spatial and temporal properties of the displays. Human performance, however, was far worse than an ideal observer model, especially for low density displays and displays with few transformation events. Constraining human performance could be spatial and temporal integration limits beyond which events cannot be combined to recover edge orientation. In sparse displays, edges may not be formed because elements are far apart and the temporal intervals between their transformations are long. Similar integration limits exist, for example, in apparent motion, where the perception of motion between two alternatively flashing elements is constrained by the inter-element distance and the element flash timing (Korte, 1915; Wertheimer, 1912). Previous work suggests that SBF performance improves with the number of frames that can be fit into a 165 ms temporal window, with additional frames adding little or no additional benefit (Shipley & Kellman, 1993), suggesting that there exists a temporal window within which events can be successfully integrated. In the event quantity condition in Experiment 1, performance improved with increasing number of events up to 25, beyond which there was no added benefit to sensitivity. This may reflect a constraint on the number of events that can be usefully integrated within a certain time interval. Additionally, both Experiment 1 and prior work demonstrate a gradual reduction in SBF perception as a function of the display's spatial and temporal properties, suggesting an effect of noise.

There are several possible sources of noise in these displays. Accurate recovery of edge orientation requires registration of element positions, spatial relations between pairs of element changes elements and times between element changes. For example, the ideal observer model showed that even slight deviations from correct temporal values could cause a 1.8° error in the orientation estimate.

Experiment 2 was designed to empirically measure sensitivity to these quantities. We did not use SBF performance in any way in these measurements, but for comparability, we used SBF-like displays (fields of randomly arranged elements). Noise measurement of low-level stimulus properties has previously been used to account for error in visual speed perception (Hürlimann, Kiper, & Carandini, 2002; Stocker & Simoncellli, 2006) and cue reliability of spatial and orientation signals in biological motion (Thurman & Lu, 2014). In SBF, low-level sources of noise would result in mis-estimation of edge orientation, resulting in poorer sensitivity. We wondered whether noise in these input variables might account for differences between human and model performance.

3.1. Method

3.1.1. Participants

The 4 participants included 3 volunteers from the University of California, Los Angeles, and one of the authors, GE. All reported having normal or corrected-to-normal vision. Two of the subjects were experienced psychophysical observers. All subjects provided informed consent and this work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

3.1.2. Design

A method of constant stimuli design was used to measure sensitivity to spatial, temporal, and angular separation between pairs of elements flashed successively. Subjects performed a two-interval forced choice (2IFC) task in which they selected the interval that contained the flashed pair of elements that were farthest apart in space (spatial separation task), farthest apart in time (temporal separation task), or which formed the smallest angle relative to horizontal (angular separation task). Below, we describe the general methods that were common to all three tasks, followed by a specific description of each task.

3.1.3. Displays

The apparatus was the same as that for Experiment 1. Stimuli consisted of a background array of 400 randomly placed white, circular elements (diameter = 0.25°) on a black background and two pairs of target elements which were identical to the background elements. All elements appeared within a 13.69° by 13.69° area centered on the middle of the screen. Each trial was composed of two intervals. In each interval, one of the pairs of elements was flashed one at a time. Three properties of the flashed elements were compared across intervals in three separate experimental sessions: spatial separation, temporal separation, and angular separation.

3.1.4. Spatial separation task

The distance between two sequentially flashed elements in a pair defined a spatial separation. Subjects compared the inter-element separations of flashed elements in the two intervals and selected the interval that contained the pair of flashed elements that were farthest apart (i.e., defined the largest inter-element distance). In one of the intervals, the spatial separation between elements was held constant; this was the reference distance. In the other – the comparison interval – the separation between elements varied. The percentage of time that the reference distance was selected as being longer than the comparison was used to define a psychometric function.

Seven reference distances were tested: 0.50, 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5 degrees of visual angle. A psychometric function was measured for each. The comparison distances used were offset from the reference distance by between $-1.24°$ and 1.24°. Ten comparison distances in that range were used for each reference and were selected to cover a range of values along the psychometric function.

Each element flashed (i.e., was invisible) for 50 ms. The second element in a pair disappeared immediately after the first element reappeared. Each pair of elements in an interval was centered on a random position within the display area. The angle formed between these elements and the horizontal was randomized across trials, but was the same for both intervals within a single trial. Whether the reference or comparison appeared first was randomized across trials. Trials with each of the seven tested reference values were intermixed. Each reference-comparison pairing was tested 20 times. In total, there were 1400 trials.

3.1.5. Temporal separation task

The time between the flashing of one element and the flashing of the second was used to define a temporal duration. As with spatial separation, one interval contained a reference duration that was held constant and the second contained a comparison duration that varied. Subjects selected the interval that contained the longest temporal separation between element flashes. The percentage of trials on which the reference duration was selected as being longer than the comparison duration was used to define a psychophysical function. Six reference durations were used: 50, 100, 150, 200, 250, and 300 ms, and a psychometric function was defined for each. Ten comparison durations were used for each reference, with offsets in the range of -180 to 180 ms. For this task only, the monitor refresh rate was set to 100 Hz to allow intervals to occur in steps of 10 ms.

Elements in a pair were separated by a distance of 3.75° and appeared in random positions of the display. The angle formed by the elements and the horizontal was randomized across trials, but was held constant between intervals in a single trial. Whether the reference appeared in the first or second interval was also randomized across trials. Trials from each of the reference durations were intermixed during the experiment. Each reference-comparison pairing was tested 20 times, creating a total of 1200 trials.

3.1.6. Angular separation task

The angle formed between two flashed elements in an interval and the horizontal defined an angular separation. Reference and comparison angular separations were shown in two intervals, and subjects selected the interval that contained the angle closest to horizontal. Five reference angles were used: 15°, 30°, 45°, 60°, and 75°. Ten comparison angles were used for each reference, with offsets in the range of $-22°$ to 22°.

As in the temporal separation task, elements within a pair were 3.75° apart and appeared in random positions in the display. The timing parameters were identical to those used in the spatial separation task. Whether the angles were positive or negative was randomized across trials. For example, on one trial the reference angle might be 30° and the comparison 10° and on another trial $-30°$ and $-10°$ respectively. In both cases, the comparison angle should be judged as closer to the horizontal. Whether the reference appeared in the first or second interval was randomized across trials. Each reference-comparison pairing was tested 20 times for a total of 1000 trials.

3.2. Procedure

Subjects sat at a distance of 89.5 cm from the monitor and had their heads stabilized by a chin-rest. Subjects were given verbal instructions that they would be making discrimination judgments between the spatial, temporal, or angular distances defined by the flashing of two dots in a field of dots. A trial began with all elements, background and target, displayed on the screen for 300 ms. A red outline of a square (7.45° by 7.45°) centered on the elements of a pair appeared for 300 ms. The square was small enough to focus attention on a particular part of the display, but large enough so that both elements making up a pair fit comfortably within it. Because the angle formed by the pair of elements was unpredictable and because the square disappeared before the first element of a pair flashed, it could not be used as a cue or reference to help determine spatial or angular separation. Pilot work had found that without this attentional cue, observers often missed the disappearance of one or both elements in a target pair since they were indistinguishable from background elements and flash durations were very short. Even with the attentional cue, it was sometimes difficult to detect element flashes. In order to prevent guessing in such cases, subjects were allowed to press a key to repeat a trial. The same reference and comparison values were retested, but a new display was generated with background and target elements appearing in new positions and with the interval order randomized.

The attentional cue remained on the screen for 300 ms and then disappeared. After a further 300 ms, the first element of the first element pair disappeared for 50 ms and reappeared. The second element in the pair then immediately disappeared for 50 ms. In the temporal separation task, a pause was inserted after the reappearance of the first element and before the disappearance of the second. This pause defined the temporal interval about which subjects made a judgment. Once the second element reappeared, all elements remained on the screen for another 300 ms, at which point the second interval began. A second attentional square was shown for 300 ms and the second pair of elements flashed one at a time. After the last target element reappeared, the display remained on the screen for another 300 ms and was then replaced by a blank, black screen. White text instructed subjects to make a response by pressing one of two keys on the keyboard to indicate whether the first or second interval contained the pair of target elements that were farthest apart (spatial task), that flashed furthest apart in time (temporal task), or that formed the smallest angle with the horizontal (angular task). If subjects missed one or more target element flashes, they were instructed to press a third key to repeat a trial. Subjects were explicitly instructed not to repeat trials in which they were unsure of the answer, but saw all four target element flashes. Subjects were given a break every 100 trials. An illustration of a trial sequence is shown in Fig. 5. The spatial, temporal, and angular tasks were run in separate sessions. Each session lasted approximately one and a half hours.

3.3. Results and discussion

3.3.1. Psychometric function fitting

Each reference and their associated 10 comparison stimuli were used to define a psychometric function depicting the percentage of time that the reference interval was perceived to contain the pair of elements that were farthest apart in space, time, or which formed the smallest angle with the horizontal as a function of the tested comparison values for each of the spatial, temporal, and angular separation tasks respectively. This resulted in seven psychometric functions per observer for the spatial task, six for the temporal, and five for the angular – one for each of the reference values used in each task. Cumulative normal distributions were fit to the data for each reference and for each subject separately using a non-linear least squares procedure. The mean and standard deviation of each function were estimated. The mean of the cumulative normal corresponds to the 50% threshold of the psychometric function. The standard deviation is inversely proportional to the slope of the psychometric function. The standard deviation can therefore be considered a measure of cue reliability (e.g., Thurman & Lu, 2014), so that larger standard deviation values (and correspondingly small slopes) indicate greater uncertainty.

3.3.2. Fitting results

For all three tasks, although there was some variability in estimated means across subjects, average means were not significantly different from 0 for any reference value for any task (*ps* > 0.05). The standard deviation estimates appear in Tables 1–3 for all subjects in each of the three tasks. Note that standard deviations should not be directly compared across tasks since the underlying units are different for each task (degrees of visual angle, milliseconds, and degrees).

The standard deviations from each condition were submitted to a one-way, repeated-measures ANOVA. There was a significant effect of reference spatial separation ($F(6,18) = 5.84$, $MSE = 0.02$, $p = 0.002$, $\eta^2_p = 0.66$) and angular separation ($F(4,12) = 4.30$, $MSE = 7.47$, $p = 0.022$, $\eta^2_p = 0.59$), but no effect of reference temporal separation ($F(5,15) = 0.17$, $MSE = 574$, $p = 0.97$, $\eta^2_p = 0.05$). Despite these differences and despite differences in standard deviations across subjects (for example, subject YX's standard deviation estimates for spatial separation (Table 1) were two to three times larger than those of the other subjects), as a first step, we sought to test as simple and general a model as possible by averaging standard deviation estimates across all references and all

Table 1
Standard deviation estimates for each subject in the spatial task for each of the seven reference distances.

Observer	0.5°	1°	1.5°	2°	2.5°	3°	3.5°
GE	0.16	0.18	0.24	0.26	0.44	0.33	0.47
RO	0.27	0.35	0.35	0.32	0.42	0.41	0.79
SC	0.20	0.20	0.19	0.29	0.25	0.37	0.35
YX	0.75	0.54	0.68	0.96	0.88	1.00	1.58
Avg.	0.35	0.32	0.36	0.46	0.50	0.53	0.80

Table 2
Standard deviation estimates for each subject in the temporal task for each of the six reference temporal durations.

Observer	50 ms	100 ms	150 ms	200 ms	250 ms	300 ms
GE	68.54	133.45	99.80	91.59	107.02	86.23
RO	48.28	27.8014	50.83	51.53	69.63	62.15
SC	116.63	61.89	67.92	72.83	96.19	120.48
YX	182.43	181.26	147.96	192.33	137.94	153.60
Avg.	103.97	101.10	91.63	102.07	102.70	105.61

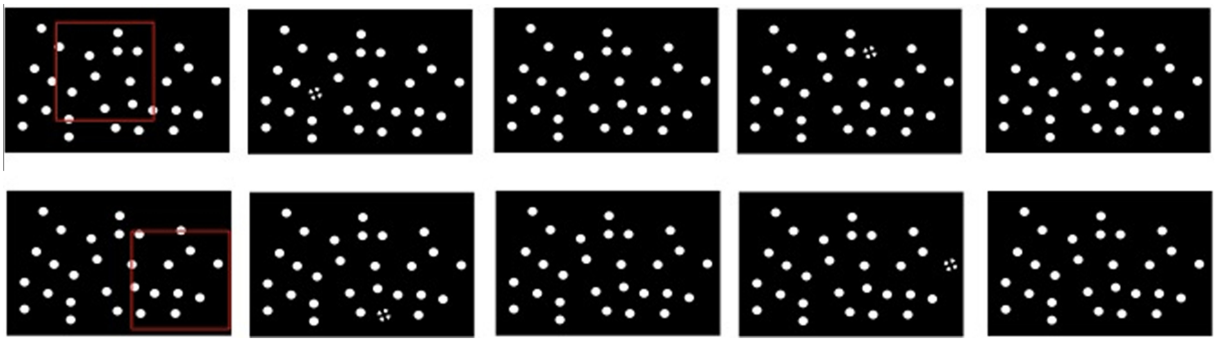

Fig. 5. An illustration of a trial in Experiment 2. Each row depicts one interval. A region of the display was cued with a red outline of a square in which element transformations would occur (first panel). An element within that region would disappear (second panel, indicated by dashed boundary) and reappear (third panel). A second element would disappear, also within the cued region (fourth panel) and reappear (fifth panel). The two elements define a spatial, angular, and temporal separation. The second row depicts a second pair of flashing elements that represent what might have been seen in the second interval. In this example, the elements in the second row are farther apart than in the first, but the angular separation between them and the horizontal is held constant across the two intervals.

Table 3
Standard deviation estimates for each subject in the angular task for each of the five reference distances.

Observer	15°	30°	45°	60°	75°
GE	6.96	7.94	12.85	9.27	7.87
RO	7.89	13.51	9.88	14.31	10.23
SC	6.14	6.41	8.17	12.13	6.610
YX	5.45	14.84	20.58	19.33	13.46
Avg.	6.61	10.67	12.87	13.76	9.54

subjects. The resulting average estimates were 0.47 degrees of visual angle, 101.18 ms, and 10.69° for spatial, temporal, and angular separation respectively. Instead of using the data from Experiment 2 to fit individual performance data, we used the averages to predict performance from a completely different group of subjects from Experiment 1.

3.3.3. Model construction and results

The ideal observer model introduced in Experiment 1 was modified by adding two constraints and three sources of noise. The first constraint was on the number of integrated elements from which the final orientation estimate was derived. In the event quantity condition in Experiment 1, threshold estimates were constant for 25 and more events; we therefore restricted the number of elements to be integrated to 25. A sequence of 25 consecutive element transformations was sampled for each trial, and orientation estimates were derived only from triplets within that set of elements. The second constraint was temporal: Triplets containing inter-event times greater than 165 ms were excluded from the final set from which the average orientation was computed. Previous work found that perception of SBF was greatly reduced beyond this limit (Shipley & Kellman, 1994).

The average noise parameters estimated in Experiment 2 were applied by including additive noise to the spatial (A), temporal (ΔT), and angular (φ) inter-element properties as indicated in Eq. (1). This produced the following revised equation:

$$\hat{\theta} = \tan^{-1}\left(\frac{\hat{v}_{23} * \sin\hat{\varphi}_{23} - \hat{v}_{12} * \sin\hat{\varphi}_{12}}{\hat{v}_{23} * \cos\hat{\varphi}_{23} - \hat{v}_{12} * \cos\hat{\varphi}_{12}}\right) \tag{3}$$

$$\hat{v}_{ij} = \frac{\hat{A}_{ij}}{\Delta\hat{T}_{ij}} \tag{4}$$

The accents above the variables in Eqs. (3) and (4) indicate that they are estimates. An example for how these estimates were computed is shown below for the angular property:

$$\hat{\varphi}_{ij} = \varphi_{ij} + N(0, \sigma^2_{\varphi}) \tag{5}$$

In Eq. (5), the true angular separation was corrupted by a noise drawn from a normal distribution with a mean of zero and a standard deviation given by the average standard deviation derived in Experiment 2 (i.e., the average across all four subjects and references). A mean of zero of was used because there was no evidence for bias in the averaged data from Experiment 2. For spatial and temporal noise, truncated normal distributions were used to ensure that the final spatial and temporal estimates were non-negative.

Each condition in Experiment 1 was simulated ten times. For each simulation, a staircase procedure was repeated for each stimulus level in each condition. On each trial, the sequence of element transformations was recorded and divided into triplets. Triplets that included temporal separations between element transformations that exceeded 165 ms were excluded. Of the remaining triplets, 23 consecutive triplets were chosen (corresponding to 25 events). (For the number of events condition, if there were fewer than 23 events, then that smaller number of events was used). For each triplet, an estimate of the bar's orientation, $\hat{\theta}$, was computed. Each time the estimate was computed, new noise samples were drawn for each of the three display parameters as indicated above. The median of the orientation estimates derived from all triplets was used as the final estimate for one trial. This single estimate was translated to a "clockwise" or "counter-clockwise" response and submitted to the adaptive staircase. The averages of the threshold estimates from the ten simulations are shown in Fig. 6. Model results are shown in blue; human performance from Exp. 1 is shown in black.

Model performance was evaluated by computing the root mean squared error between the human orientation discrimination thresholds and the model's. The model matched human performance very well across all conditions: element density RMSE = 3.42°; number of events RMSE = 0.89°; frame duration RMSE = 0.83°. It is important to note that the same noise parameters estimated from Experiment 2 were used for simulating model thresholds for all three conditions and that the model was producing orientation estimates on a triplet-by-triplet basis, while the comparison between model and human performance was at the level of orientation discrimination thresholds.

In addition to this model, we examined several alternatives. In the spatial separation task in Experiment 2, cue reliability decreased (standard deviations increased) as a function of reference distance. That is, estimates of inter-element distance were more variable for larger separations than for smaller ones. In the implementation of the model, we chose to ignore this variation

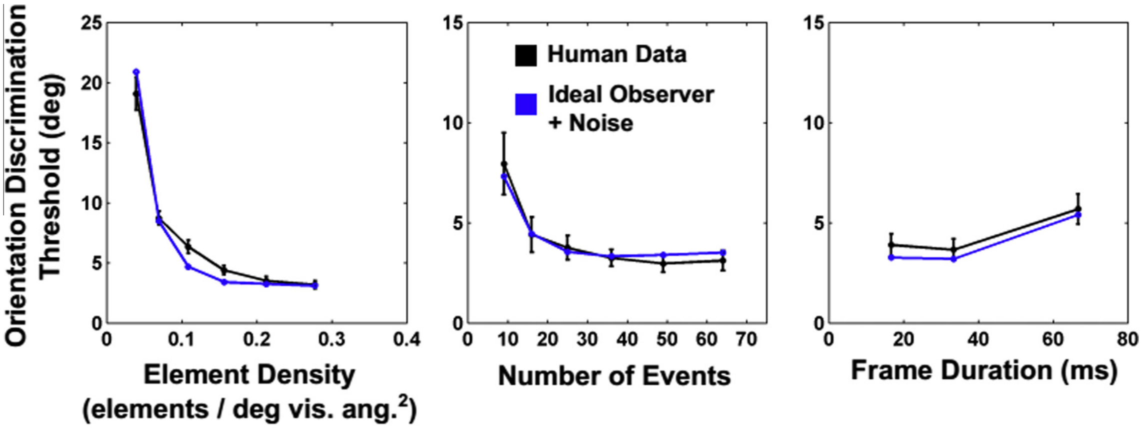

Fig. 6. Data from Experiment 1 (black) replotted with model fits (blue) using noise parameter estimates from Experiment 2. Error bars indicate ± standard error of the mean. Model data reflect the average of 10 simulated experiment runs. Note the difference in scale of the y-axis between the first and second and third graphs.

and used additive noise which added the same amount of noise irrespective of inter-element separation. However, increasing variability in distance perception as a function of inter-element separation suggests that spatial noise may be multiplicative in nature rather than additive. A role for multiplicative computations has previously been suggested for looming signals (Gabbiani et al., 2002), contrast-gain control (Albrecht & Geisler, 1991; Määttänen & Koenderink, 1991), and orientation selectivity (Beaudot & Mullen, 2005). To test whether multiplicative noise might better capture human performance, the spatial separation data from Experiment 2 were log-transformed and refit with a cumulative normal, and the standard deviation for each distance for each subject was recomputed. The average was again taken across all references and subjects, resulting in a new spatial noise parameter of 0.30. The simulations were then repeated for all conditions. The multiplicative noise model was able to fit the density condition slightly better than the additive noise model (RMSE = 3.11°), but was worse for the number of events (RMSE = 1.74°) and frame duration (RMSE = 2.15°) conditions.

We also looked at two simplified versions of the model that included only either the additive noise for spatial, temporal, and angular separation (noise-only model), or the two constraints on number of integrated events and on the time between element transformation events (constraints-only model). The goal of this comparison was to see whether one or the other set of parameters could account for the human data on its own. Model thresholds were computed only for the density condition. Both the noise-only and constraints-only models produced much worse fits to the human data with RMSEs of 13.15° and 11.18° respectively.

It was surprising that a model with a few simple noise parameters and constraints was able to simulate human performance so well despite the fact that the noise parameters were derived from data averaged across subjects and conditions and despite the fact that those subjects did not participate in the first experiment. It is also important to note that this model had no free parameters because the three noise parameters and two constraints were determined empirically from separate experiments. This suggests that the model was successfully capturing a perceptual process.

4. General discussion

The current studies support a model of SBF that extracts local, oriented edge fragments from small groups of element changes. We obtained experimental data on SBF for a single edge orientation under variations of element density, element number, and frame duration, all of which substantially affected performance in an objective discrimination task. These experiments, and some prior work, also suggested two constraints limiting information integration in SBF. One constraint is that, at least as tested here, SBF performance improved up to about 25 change events and not beyond. The other, obtained from prior work, is a limitation of integration to a 165 ms temporal window. Separately, using non-SBF tasks, we obtained experimental data on the variability of encoding basic inputs to SBF, specifically element separation, relative angle of event pairs, and temporal separation of events.

These results were sought in order to understand the processing underlying SBF, specifically to incorporate in a model realistic estimates of noise in human encoding of the inputs. The model describes a spatiotemporal edge perception mechanism that uses relationships of sequential, discrete element changes to establish a local, oriented edge fragment.

Data from Experiment 1 were compared to an ideal observer version of the model that samples edge triplets and uses information optimally. That model performed the task of Experiment 1 to near perfection from as little as the theoretical minimum of one non-collinear triplet of sequential change events. This output well exceeded human performance. We used the experimental findings in Experiments 1 and 2 directly, however, to build a more realistic model. Limiting accumulation of change events to no more than 25 and limiting temporal integration to a 165 ms interval were the key spatial and temporal constraints implemented in the revised model. In addition, noise estimates obtained experimentally were used to create noise distributions for spatial separation, angular relation, and temporal separation. The revised SBF process model was configured to sample from displays and noise distributions and, like human participants, to give trial-by-trial responses. As with human participants, the trial-by-trial data were inputs to staircase procedures that determined orientation discrimination thresholds.

The modeling results showed remarkably good fits to the human data, with no free parameters. Three very different manipulations – element number, element density, and frame duration – were all fit closely with the same model parameters. These results are consistent with the idea that the foundation of SBF is a spatiotemporal edge perception mechanism that works as modeled here. As noted earlier, for perception of global form in more complex shapes, an additional stage in which local edge fragments are connected by well-known interpolation processes would produce the rich shape perception phenomena shown in earlier SBF research (Erlikhman, Xing, & Kellman, 2014; Shipley & Kellman, 1994).

There are several limitations of the current modeling effort. First, the model used noise distributions for each of the noise parameters tested in Experiment 2 that were averaged over several participants. We also tested the model by the fit to data averaged over participants in the SBF experiments (Exp.1). It could have been the case that individual variations were too extreme to obtain good fits using these simplifications, but the results suggest otherwise. Second, although we obtained noise distributions at each value tested in the noise experiments (5 angular relations, 6 temporal separations, and 7 spatial separations), we incorporated into the model a single noise distribution for each of these three noise variables, obtained in each case by taking the average distribution across the several values tested for each variable. Looking at the variation across values, this simplification was most clearly justified for temporal separation, where the several values tested yielded highly similar results. It was somewhat less justified for spatial separation, which gave some indication of a monotonic increase in variability from the lowest to highest values testes, and for angular relations, which produced some evidence for more precise performance at the values nearest 0 (vertical) and 90 degrees (horizontal), and somewhat less precise performance in between. These simplifications of averaging noise estimates over participants and fitting group data was used in order to provide a basic, initial test of the model. The error bars on the participant data give some indication of the variability for human participants, and in almost all cases, the model predictions fall squarely within ± 1 standard error of the mean for human participants. The averaging over values on the noise dimensions tested seems unavoidable, as stimuli for actual SBF displays will include a variety of element change triplets that vary in spatial separation, temporal separation, and angular relations.

There may be several limitations to the applicability of the model. First, the calculations in the model that extract a local edge orientation estimate from triplets of changes assume that the velocity and orientation of the virtual edge is constant. We have recently shown that SBF supports a wide range of transformations of the virtual object including scaling, rotation, acceleration, and non-rigid transformation (Erlikhman, Xing, & Kellman, 2014). It is not yet clear how these findings relate to the current model. The recent findings suggest that the local edge extraction stage is fast and can occur with relatively few element changes. If so, the current model may indeed work for rigid and non-rigid shape changes, but there may be predictable ways in which edge and shape recovery break down when shape, orientation, or scale changes occur too rapidly. Testing the current model in more complex displays is an important priority for further research. Second, the current model only gives the orientation for a single edge. In displays with 2D shapes or curved edges, the model would need some spatial parameter that limits the integration of element transformations to small neighborhoods to allow for the extraction of different edge positions and orientations along different portions of the contour. To give an extreme example, if the illusory figure is a circle, then there must be some way of keeping separate element transformations that occur on opposite sides of the circle. One possibility is that orientation and velocity are treated as approximately constant within a small spatiotemporal integration window. Edge fragments can then be recovered simultaneously within several such windows around the boundary of the object. Future work is needed to address these issues in detail.

For a moving object, two frames from a motion sequence are sufficient to solve an aperture problem that occurs locally for each contour (e.g., Weiss, Simoncelli, & Adelson, 2002). However, as usually tested, the contours are given by oriented contrast in static views; thus, the solution to the aperture problem gives a motion direction. In SBF, no oriented contrast edge is given in any static view. This more complicated task of recovering both edge orientation and local edge motion in SBF has been called the *point aperture* problem (Prophet, Hoffman, & Cicerone, 2001).

Although the point aperture problem involves recovery of orientation and motion, whereas the classical aperture problem involves recovering motion with orientation already given, an interesting conjecture is that the spatiotemporal edge perception mechanism at the foundation of SBF relies on the same spatiotemporal filters that have been theorized to underlie motion perception. Motion energy models describe spatiotemporal filters embodied in basic neural mechanisms which can detect moving, contrast-defined edges over time (Adelson & Bergen, 1985; Challinor & Mather, 2010; van Santen & Sperling, 1984). Could these filters serve as spatiotemporal edge detectors when oriented edges are not given by contrast (or other static spatial information)? An apparent problem with this conjecture is that motion energy models applied to our displays would predict nearest-neighbor apparent motion between transforming elements, which is not what is seen in SBF (Shipley & Kellman, 1994). A modified version of this conjecture is that a set of large, oriented spatiotemporal filters that capture within their receptive fields the transformations of several elements may be used to determine edge orientation. Evidence for the existence of such filters have been found in primates in V1 (Marcar et al., 2000; Schmid, 2008), V2 (Chen et al., 2014; Lu et al., 2010) and MT (Marcar & Cowey, 1992; Marcar et al., 1995). There are a number of ambiguity problems that would have to be resolved to explain how responses of multiple detectors at different orientations and scales converge on unified edge fragment orientation and local motion, but this is an interesting possibility, worthy of detailed evaluation. If an explanation of the first stage of SBF in terms of spatiotemporal orientation/motion filters is possible, then SBF is not simply an esoteric visual illusion, but is at bottom the result of a fundamental visual process involved in the extraction of edges, motion, and their interactions. We are currently exploring the possibility of linking models of SBF to the known properties of oriented motion energy filters, an enterprise that could be especially fruitful in elucidating important relations between basic visual filtering and high-level phenomena of perceptual organization.

Acknowledgments

We would like to thank Maxwell Mansolf, Renato Fillinich, and Ariela Iancu for greatly contributing the early work on this project. Also, we thank Hongjing Lu, Dario Ringach, Greg Francis, and James Pomerantz for helpful comments.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perceptions of motion. *Journal of the Optical Society of America A, 2*(2), 284–299.

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300*, 523–525.

Albrecht, D. G., Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in visual cortex.

Andersen, G. J., & Cortese, J. M. (1989). 2-D contour perception from kinematic occlusion. *Perception and Psychophysics, 46*, 49–55.

Barraza, J. F., & Chen, V. J. (2006). Vernier acuity of illusory contours defined by motion. *Journal of Vision, 6*, 923–932.

Beaudot, W. H. A., & Mullen, K. T. (2005). Orientation selectivity in luminance and color vision assessed using 2-d band-pass filtered spatial noise. *Vision Research, 45*(6), 687–696.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Challinor, K. L., & Mather, G. (2010). A motion-energy model predicts the direction discrimination and MAE duration of two-stroke apparent motion at high and low retinal luminance. *Vision Research, 50*, 1109–1116.

Chen, M., Li, P., Zhu, S., Han, C., Xu, H., Fang, Y., Hu, J., Roe, A. W., Lu, H. D. (2014). An orientation map for motion boundaries in V2. Cerebral Cortex, bhu235.

Cicerone, C. M., Hoffman, D. D., Gowdy, P. D., & Kim, J. S. (1995). The perception of color from motion. *Perception & Psychophysics, 57*, 761–777.

Cooke, T., Cunningham, D. W., & Bülthoff, H. H. (2004). The perceptual influence of spatiotemporal noise on the reconstruction of shape from dynamic occlusion. *Lecture Notes in Computer Science, 3175*, 407–414.

Cropper, S. J. (2005). The detection of motion in chromatic stimuli: First-order and second-order spatial structure. *Vision Research, 45*, 865–880.

Cropper, S. J., & Derrington, A. M. (1994). Motion of chromatic stimuli: First-order or second-order? *Vision Research, 34*(1), 49–58.

Cunningham, D. W., Shipley, T. F., & Kellman, P. J. (1998). Interactions between spatial and spatiotemporal information in spatiotemporal boundary formation. *Perception & Psychophysics, 60*(5), 839–851.

Derrington, A. M., & Henning, G. B. (1993). Detecting and discriminating the direction of motion of luminance and colour gratings. *Vision Research, 33*(5/6), 799–811.

Erlikhman, G., Xing, Y. Z., & Kellman, P. J. (2014). Non-rigid illusory contours and global shape transformations defined by spatiotemporal boundary formation. *Frontiers in human neuroscience, 8*.

Fantoni, C., & Gerbino, W. (2003). Contour interpolation by vector-field combination. *Journal of Vision, 3*, 281–303.

Gabbiani, F., Krapp, H. G., Koch, C., & Laurent, G. (2002). Multiplicative computation in a visual neuron sensitive to looming. *Nature, 420*, 320–324.

Gibson, J. J., Kaplan, G. A., Reynolds, H. N., & Wheeler, K. (1969). The change from visible to invisible: A study of optical transitions. *Perception and Psychophysics, 3*, 113–116.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92*, 173–211.

Hürlimann, F., Kiper, D. C., & Carandini, M. (2002). Testing the Bayesian model of perceived speed. *Vision Research, 42*, 2253–2257.

Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.

Kaplan, G. A. (1969). Kinetic disruption of optical texture: The perception of depth at an edge. *Perception & Psychophysics, 6*, 193–198.

Kellman, P. J., Erlikhman, G., Mansolf, M., Fillinich, R., & Iancu, A. (2012). Modeling spatiotemporal boundary formation. *Journal of Vision, 12*(9), 881. http://dx.doi.org/10.1167/12.9.881.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23*, 141–221.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research, 39*(16), 2729–2737.

Korte, A. (1915). Kinematoskopishe Untersuchungen. *Zeitschrift für Psychologie mit Zeitschrift für angewandte Psychologie, 72*, 194–296.

Lu, H. D., Chen, G., Tanigawa, H., & Roe, A. W. (2010). A motion direction map in macaque V2. *Neuron, 68*, 1002–1013.

Määttänen, L. M., & Koenderink (1991). Contrast adaptation and contrast gain control. *Experimental Brain Research, 87*, 205–212.

Marcar, V. L., & Cowey, A. (1992). The effect of removing superior temporal cortical motion areas in the macaque monkey: II. Motion discrimination using random dot displays. *European Journal of Neuroscience, 4*, 1228–1238.

Marcar, V. L., Raiguel, S. E., Xiao, D., & Orban, G. A. (2000). Processing of kinetically defined boundaries in areas V1 and V2 of Macaque monkey. *Journal of Neurophysiology, 84*(6), 2786–2798.

Marcar, V. L., Xiao, D. K., Raiguel, S. E., Maes, H., & Orban, G. A. (1995). Processing of kinetically defined boundaries in the cortical motion area MT of the macaque monkey. *Journal of Neurophysiology, 74*, 1258–1270.

Michotte, A., Thines, G., & Crabbe, G. (1964). *Les complements amodaux des structures percpectives. Studia Psychologica*. Louvain: Publications Unvisersitaires de Louvain.

Miyahara, E., & Cicerone, C. M. (1997). Color from motion: Separate contributions of chromaticity and luminance. *Perception, 26*, 1381–1396.

Nakayama, K., & Silverman, G. H. (1988a). The aperture problem—I. Perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Research, 28*(6), 739–746.

Nakayama, K., & Silverman, G. H. (1988b). The aperture problem—II. Spatial integration of velocity information along contours. *Vision Research, 28*(6), 747–753.

Palmer, E. M., Kellman, P. J., & Shipley, T. F. (2006). A theory of dynamic occluded and illusory object perception. *Journal of Experimental Psychology: General, 135*(4), 513–541.

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

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. <http://www.palamedestoolbox.org>.

Prophet, W. D., Hoffman, D. D., & Cicerone, C. M. (2001). Contours from apparent motion: A computational theory. In P. Kellman & T. Shipley (Eds.), *From fragment to objects: Segmentation and grouping in vision* (pp. 509–530). Amsterdam: Elsevier Science Press.

Schmid, A. M. (2008). The processing of feature discontinuities for different cue types in primary visual cortex. *Brain Research, 1238*(31), 59–74.

Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research, 29*(5), 619–626.

Shipley, T. F., & Kellman, P. J. (1993). Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion? *Spatial Vision, 7*(3), 323–339.

Shipley, T. F., & Kellman, P. J. (1994). Saptiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology: General, 123*, 3–20.

Shipley, T. F., & Kellman, P. J. (1997). Spatio-temporal boundary formation: The role of local motion signals in boundary perception. *Vision Research, 27*(10), 1281–1293.

Stappers, P. J. (1989). Forms can be recognized from dynamic occlusion alone. *Perceptual and Motor Skills, 68*, 243–251.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*(4), 578–585.

Thurman, S. M., & Lu, H. (2014). Bayesian integration of position and orientation cues in perception of biological and non-biological forms. *Frontiers in Human Neuroscience, 8*, 91. http://dx.doi.org/10.3389/fnhum.2014.00091.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, Massachusetts: M.I.T. Press.

van Santen, J. P. H., & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A, 1*, 451–473.

Wallach, H. (1935). Uber visuell wahrgenommene bewegungsrichtung. *Psychologische Forschung, 20*, 325–380.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598–604.

Wertheimer, M. (1912). Experimentelle Studien über das Segen von Bewegung. *Zeitschrift für Psychologie, 61*, 161–265.

Yonas, A., Craton, L. G., & Thompson, W. B. (1987). Relative motion: Kinetic information for the order of depth at an edge. *Perception & Psychophysics, 41*, 53–59.