VISION RESEARCH

An International Journal for Functional Aspects of Vision

Classification images reveal spatiotemporal contour interpolation ☆

Brian P. Keane [a,*], Hongjing Lu [b], Philip J. Kellman [a]

[a] UCLA Human Perception Laboratory and Psychology Department, University of California, Los Angeles,
1285 Franz Hall, Los Angeles, CA 90095-1563, USA
[b] Psychology Department, University of Hong Kong, Hong Kong

Abstract

Contour interpolation is the process whereby spatially separated object fragments (inducers) are connected on the basis of their contour relations. An important characteristic of interpolation between simultaneously presented inducers is that observers rely on interpolation regions to perform a discrimination task. However, it is unclear if the same property holds when inducers are separated in both space and time. To address this question of *spatiotemporal interpolation*, we had participants discriminate spatiotemporally presented "fat" and "thin" noise-corrupted figures, when the figures were stationary (Experiment 1) or moving (Experiment 2), and when the connections across vertical gaps were either real, interpolated (illusory), or absent. Classification images from both experiments showed that noise regions near interpolated boundaries affect performance comparably to when real contours appear, but very little in the absence of interpolation. The classification images also revealed information about the time course of interpolation and suggested that contour interpolation between simultaneously visible inducers may be a special case of a more general spatiotemporal contour interpolation process.[1]

© 2007 Published by Elsevier Ltd.

1. Introduction

Being that objects in ordinary visual scenes are partially occluded, an important priority of visual processing is to determine which visible fragments belong to the same object. Central to this task is *contour interpolation*, the process whereby spatially separated object fragments are connected on the basis of their edge relations. Interpolation aids in the determination of, among other things, how many objects are seen at a time and what shapes those objects have. As a canonical example, the Kanizsa square is viewed, not as notched circles, but as four complete circles behind a single square that is partly camouflaged by the background.

There are a number of models describing the conditions under which interpolation occurs in static arrays (for reviews, see Fantoni & Gerbino, 2003; Kellman, Guttman, & Wickens, 2001). According to the model of Kellman and Shipley (1991), contour interpolation occurs between visible contours (hereafter, *inducers*) that are geometrically relatable, where relatable inducers are those that can be connected by a smooth, monotonic curve that bends no more than about 90 deg. Interpolation, according to this model, typically occurs between first-order tangent

[1] The following acronyms appear in this paper: RMS—root mean square; SNR—signal to noise ratio; CI—Classification image.

discontinuities (Rubin, 2001; Shipley & Kellman, 1990; but see Tse & Albert, 1998), which are sharp corners or junctions of visible fragments.

Although the capacity to recover the shape and number of objects from fragmentary information has most often been studied with static displays, the visual system must cope with fragmentation in space and time. When the perceiver, an object, or both are moving, different parts of the same object may project to the eyes at different times and some parts of the object may never project. Notwithstanding the challenges presented by this spatiotemporal fragmentation, perceivers appear to have impressive capabilities for representing objects as coherent and persisting (Bruno & Bertamini, 1990; Kellman & Cohen, 1984; Palmer, Kellman, & Shipley, 2006). An object moving swiftly behind foliage will be represented not as a series of disconnected parts, but as a single, persisting thing. Recent evidence suggests that this ability depends heavily on processes that collect and interpolate contours between fragments that appear sequentially over time. This process we refer to as *spatiotemporal contour interpolation* and will be the focus of the present paper.

1.1. Spatiotemporal contour interpolation

Many studies have explored how the visual system represents shape from information given over time (e.g., Barenholtz & Feldman, 2006; Bruno & Gerbino, 1991; Burr, 1979; Helmholtz, 1867/1962; Kandil & Lappe, 2007; Morgan, Findlay, & Watt, 1982; Nishida, 2004; Parks, 1965; Burr & Ross, 1986). Fewer studies have looked at spatiotemporal contour interpolation, in which parts of a shape are never physically specified, and where contour interpolation must fill-in the missing boundaries.[2] In the present work, we study spatiotemporal illusory contour formation, where the elements that induce a given interpolated contour never simultaneously appear.

In possibly the first study of spatiotemporal contour interpolation (Kellman & Cohen, 1984), a black illusory figure was induced by sequential interruptions in spatially separated white elements on a black background. Displays were arranged so that the illusory form could not be seen from single or multiple static frames. The kinetic illusory figures were seen when either the virtual figure rotated or the background rotated. These results, along with others (Bruno & Bertamini, 1990), indicate that inducing events separated in both space and time can produce perception of complete objects, although no formal model of the process was proposed.

Palmer and colleagues (2006) carried out a more comprehensive analysis of spatiotemporal contour interpolation. They proposed that the connection of object fragments across space and time is governed by *spatiotemporal relatability*. They hypothesized that the processing of spatiotemporal relatability was carried out via a dynamic visual icon (DVI) which involves (1) persistence of recently viewed, but now occluded or camouflaged fragments (Neisser, 1967; Sperling, 1960), and (2) a position updating mechanism, which predicts changes in the position of a recently viewed object fragment (see Fig. 1). These processes of persistence and position updating allow a determination as to whether recently and currently viewed fragments satisfy the spatial geometric constraints of relatability. The mechanisms of the DVI and relatability, operating together, offer a possible account of how objects are coherently perceived despite spatiotemporal interruptions in their appearance.

Spatiotemporal relatability as an account of spatiotemporal contour interpolation was tested by Palmer et al. (2006) in a series of experiments. On each trial, subjects saw three object fragments moving together behind apertures. Subjects were subsequently presented with two simultaneously presented arrays—perfectly aligned fragments and slightly misaligned fragments (see Fig. 2). The task was to determine which set of fragments they had just observed behind the apertures. On the basis of spatiotemporal relatability, the authors predicted that when a fragment set could be formed into a complete object via spatiotemporal contour interpolation, discrimination would be more accurate relative to when a set was unrelatable or when a set had rounded corners (no tangent discontinuities). The predictions were confirmed: spatiotemporal
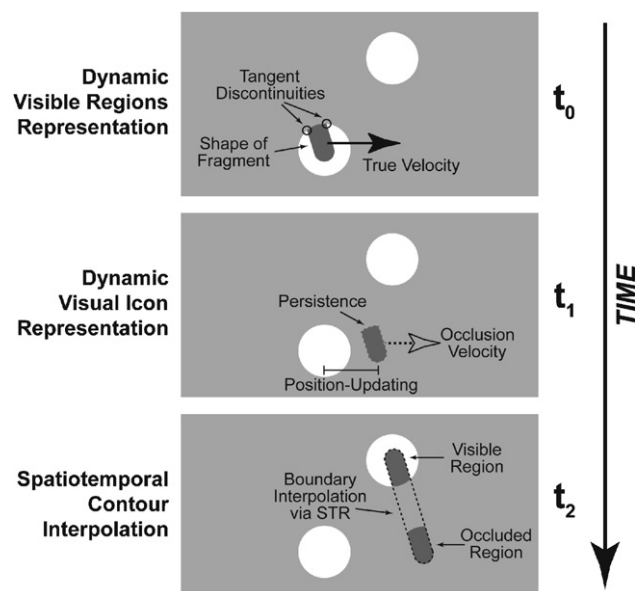
Fig. 1. A model of the spatiotemporal contour interpolation process. At time t_0 a rod fragment becomes visible; at time t_1 that entire rod becomes invisible; and, finally, at time t_2 the top part of the rod appears. In order to interpolate between the two rod portions, Palmer et al. (2006) suggest that a dynamic visual icon stores contour and velocity information of the fragment appearing at t_0, and utilizes this information to interpolate with the fragment appearing at t_2. (Figure derived from Palmer et al., 2006, p. 537.)

---

[2] For brevity, we will use the terms "spatiotemporal interpolation" or simply "interpolation" to refer to spatiotemporal contour interpolation, unless clarity requires otherwise.
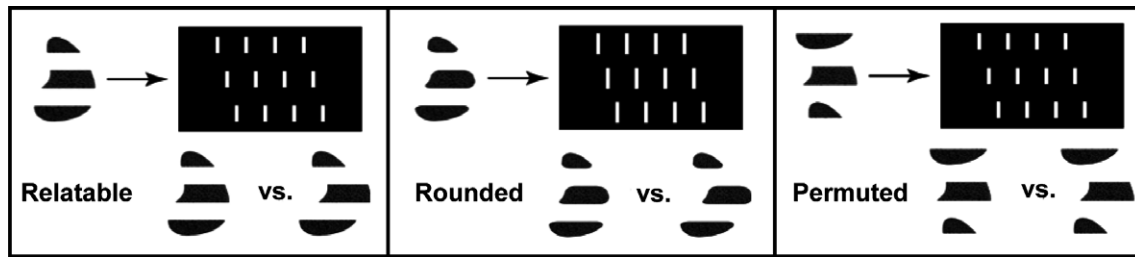
Fig. 2. A paradigm for examining spatiotemporal contour interpolation. In Palmer et al. (2006), subjects discriminated vertically aligned and misaligned sets of object fragments that moved behind a holed occluder. Discriminations were best when the fragments of the aligned set were relatable, to form a coherent figure; second-best when the corners of the relatable condition were smoothed ("rounded" condition), so as to weaken interpolation; and worst when the top and bottom fragments of the relatable condition were swapped (permuted) so that interpolation was completely eliminated. These results along with others suggest that even though different pieces of an object appear at different times, the visual system can interpolate between those pieces to engage in unit formation. (Figure adapted from Palmer et al., 2006, p. 519.)

relatability led to markedly higher sensitivity to the alignment relations of fragments, and the presence of tangent discontinuities contributed towards the sensitivity. Analogous to numerous findings with stationary displays, these data were interpreted as owing to the formation of unitary objects, when viewed fragments were relatable. These results, along with a number of direct comparisons of dynamic interpolation with static controls, supported the basic proposals of persistence, position updating, and the application of relatability in spatiotemporal interpolation.

1.2. Filling-in during contour interpolation

Contour interpolation involves representing stimulus regions as having contours when in fact no contours are visible. What are the functional consequences of this filling-in process? To what extent, if any, can information from interpolation regions affect perceptual performance? An ideal observer would ignore task-irrelevant parts of a stimulus, but experimental results suggest that human observers in contour interpolation tasks and other filling-in tasks (Nishida, 2004; Watamaniuk & McKee, 1995; Yantis & Nakama, 1998) are not ideal in this respect. For example, when subjects repeatedly discriminated "fat" and "thin" Kanizsa squares (Ringach & Shapley, 1996), performance became worse when fixed, straight-line segments were presented near the interpolation paths. Crucially, when the same lines appeared, but the inducers no longer formed a fat/thin square (so that interpolation was absent), the fixed lines did not affect performance.

Gold, Murray, Bennett, and Sekuler (2000) confirmed Ringach and Shapley's finding in a response discrimination or *classification image* (CI) paradigm (Ahumada, 1996; Ahumada & Lovell, 1971). In that study, subjects repeatedly discriminated fat or thin figures, which were corrupted with static luminance noise. In three of five conditions, those figures were real squares (real condition), interpolated Kanizsa squares (illusory condition), or fragmented squares (fragmented condition). The classification image technique, to be discussed further below, revealed correlations between pixel luminance and observer response and showed that: (a) noise pixels near interpolated contours

in the illusory condition influenced observer response, even though those pixels were objectively task-irrelevant; (b) the degree of influence was comparable to when real contours were present along those same paths; and (c) subjects were responsive primarily to regions near the visible contour fragments in the absence of real or interpolated boundaries (the "fragmented" condition, see Fig. 3).

Gold and Shubel (2006) extended the CI technique to investigate temporal properties of spatial contour interpolation.[3] The experiment was similar to that of Gold et al. (2000) except that there were two conditions (real and illusory) and the figures were shown in dynamic luminance noise. A new CI was computed for each frame of the dynamic noise, and the resulting classification image "movie" showed that pixels near illusory contours became increasingly influential across the first 175 ms of stimulus presentation.

1.3. Methods and motivations

In our experiments, we also employed classification images to study contour interpolation. The CI methodology draws from signal detection theory, and models an observer's response as being based on a decision variable s, which equals the cross-correlation of a template T and a stimulus I (Murray, Bennett, & Sekuler, 2002, p. 79; Green & Swets, 1966). When I is corrupted with additive external Gaussian white noise N, and when N is much greater than co-existing internal noise, internal noise can be neglected, and s can be expressed as:

$$s = (I + N) \otimes T$$

In a two-alternative task, the observer (assumed to be a linear discriminator) will set some criterion c, and then give one response if $s \geqslant c$, and give the alternative response otherwise. The response classification methodology provides an estimate for T by indicating the degree of influence

---

[3] For the purposes of this paper, "spatial contour interpolation" denotes contour interpolation that spans gaps only in space.
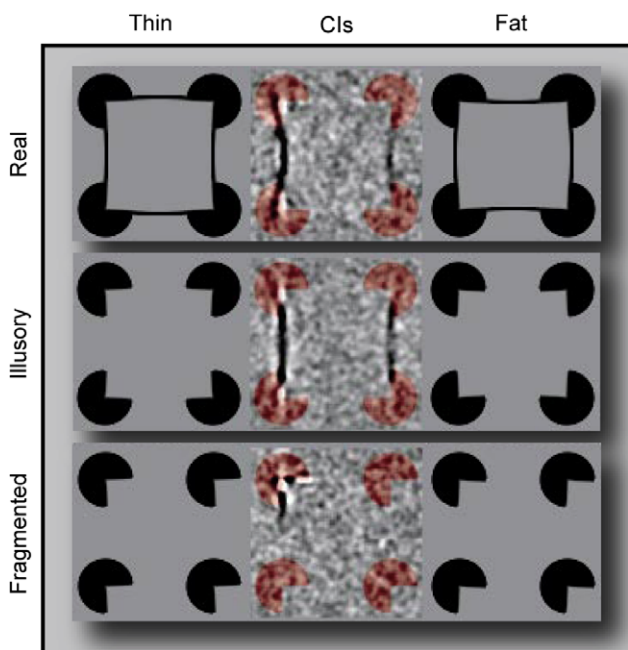
Fig. 3. Classification images revealing spatial contour interpolation (Gold et al., 2000). Each row shows for a condition, high contrast alternatives (fat/thin) and an average classification image resulting from 30,000 discriminations between those alternatives. Discriminations were difficult because the alternatives were noise-corrupted and presented with reduced contrast. The CIs show that subjects were affected by information along illusory contours similarly to when actual information appeared along those contours; but when there were no shape contours (fragmented condition), noise pixels near the inducers were primarily influential. (Figure derived from Gold et al., 2000, p. 663.)

each pixel has on the observer. A CI reveals—in picture form—these pixel influences, with darkest and lightest pixels exercising the greatest effect.[4]

In the present paper, we use CIs to examine contour interpolation between sequentially appearing inducers. The most basic question was: When inducers are separated in space and time, are interpolation regions relevant for discrimination? To address this question, subjects in two experiments discriminated figures, the tops and bottom of which were connected by luminance-defined contours (real condition), interpolated contours (illusory condition), or no contours at all (fragmented condition). In Experiment 1, the figures were stationary and were embedded in static luminance noise. These figures became visible by gradually occluding and disoccluding dark background elements. In Experiment 2, figures moved and were embedded in dynamic luminance noise (frame rate = 17 Hz). These fig-

ures became visible by gradually occluding dark stationary background elements. In both experiments, the spacing of the background elements ensured that different figure parts would become visible at different points in time. Consequently, in both experiments, representing the entirety of a discriminated (vertical) contour in the real and illusory conditions required accumulating information over time. In Experiment 2, representing an entire real or illusory contour additionally required updating the positions of temporarily camouflaged figure fragments. In both experiments, if CIs show that pixels near interpolated and real contours have similar influence, and that pixels between inducers do not strongly affect performance in the absence of contour formation, then spatiotemporal contour interpolation can be considered to have functional effects that go beyond a mere phenomenal presence.

A secondary aim of the present paper was to consider the relation between spatial and spatiotemporal interpolation. If CIs show that subjects are not sensitive to interpolated regions, then that would suggest that spatial and spatiotemporal contour interpolation involve fundamentally different processes. On the other hand, if our results mirror the fragmented, illusory, and real CIs of Gold et al. (2000), it would support the conjecture that spatial contour interpolation is a limiting case of spatiotemporal contour interpolation (Palmer et al., 2006). On that view, spatial interpolation may be a case of spatiotemporal interpolation that involves minimal persistence and zero position updating.

Although the goal of this study was not to evaluate time course, Experiment 2 provided insights into the microgenesis of spatiotemporal contour interpolation. In the following, we will briefly remark on how spatiotemporal contour formation unfolds over time, and how it compares with the processes involved in representing real or fragmented spatiotemporal figures.

2. Experiment 1: Spatiotemporal contour interpolation for static objects

2.1. Methods

2.1.1. Subjects

The first author and two paid UCLA students performed between 1200 and 2500 trials per session and (typically) one session/day over the course of six weeks. One of the paid subjects did not finish. To ensure external validity, 12 additional subjects (four/condition) performed two 1-h sessions in exchange for class credit. Of the 14 subjects who finished, all had normal or corrected-to-normal vision and all but the author were naive to the purposes of the experiment.

2.1.2. Apparatus

The displays were achromatic and were presented on a Macintosh computer monitor with a resolution of 1600×1200 pixels and a refresh rate of 85 Hz. Subjects

---

[4] Templates reveal a process of linear summation, as described above, but there does not appear to be a consensus as to whether they also reveal *representations*. Because extant research tentatively endorses (Gold & Shubel, 2006) or is consistent with (e.g., Abbey & Eckstein, 2002; Neri & Levi, 2006) the view that CIs reveal representations, and because we are unaware of any argument that explicitly rejects this view, we will regard templates as plausibly revealing characteristics of contour interpolation representations.

were seated in a darkened room and were positioned in a headrest 30 in. from the monitor, creating a viewable screen that subtended an angle of 29 by 22 deg. The background luminance was 37 cd/m². All displays were programmed in Matlab using Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

2.1.3. Stimuli

2.1.3.1. Oval characteristics. Eight dark ovals (four on bottom, four on top) translated horizontally from left to right at 8.4 deg/s for 447 ms. The ovals moved under an opaque figure that was the same color as the background. Numerous small ovals were chosen rather than two large ovals, because spatiotemporal interpolation has been found to be more robust in the first case (Palmer, 2003). Oval contrast determined the difficulty level (see below). The opaque figure's oblique and horizontal edges became visible only when they occluded the ovals in particular frames (see Fig. 4C). Ovals of the same row were separated from one another by 30 arcmin, and the first oval of the top and bottom rows were always misaligned by 15 arcmin. The misalignment ensured that vertically aligned inducers never provided information along a vertical edge simultaneously. In particular, the time interval between when subsequent edges became fully visible was between 12 and 24 ms. In the illusory and real conditions, the oval dimensions were 39 × 10 arcmin and the centers of the translating ovals struck the visible horizontal edges of the stationary figure. In the fragmented condition, the ovals were positioned and shaped the same as before except that the top of the lower ovals and the bottom of the upper ovals were each extended vertically by 10 arcmin. We extended the ovals in this way because the 6 arcmin misalignment between each pair of oblique edges from the top and bottom fragments in the fragmented condition was, phenomenologically, not enough to block interpolation (interpolation has been shown to tolerate up to about 20 arcmin of misalignment, Shipley & Kellman, 1992). An advantage to our non-interpolation control is that, in contrast to what some others have used (e.g., Lee & Nguyen, 2001), the vertical distance between the oblique edges of inducers in the fragmented display is the same as that of the illusory displays.

2.1.3.2. Figure characteristics. Although the stimuli in the illusory and real conditions appeared to be single rectangles, in all conditions figures were composed of two vertically aligned trapezoids that blended in perfectly with the gray background. In the illusory and real conditions, each trapezoid of a pair either tapered inward toward the center or extended slightly outward toward the center to provide the appearance of a single fat or thin rectangle. Also, only the peripheral horizontal contour of a trapezoid was visible (since the interior horizontal contour blended in with the background). In the fragmented condition—because the ovals were elongated—one observed both horizontal edges

of each trapezoid, and the trapezoids appeared either thin (tapering inward) or fat (see Fig. 4A).

In all conditions: the vertical length from the bottom of the lower trapezoid to the top of the upper trapezoid was 78 arcmin, the top of the upper trapezoid always measured 39 arcmin horizontally, and the bottom of the upper trapezoid measured 35 arcmin horizontally for the thin response and 43 arcmin horizontally for the fat response. The top trapezoid measured 19 arcmin vertically. The fragmented condition differed from the illusory condition on the lower trapezoid dimensions. In the fragmented condition, for both response types, the lower trapezoid always had the same dimensions as the top trapezoid.

In the illusory conditions, the lower trapezoid was a reflection about the horizontal axis of the upper trapezoid. The real condition was exactly the same as the illusory condition except that the former additionally involved real contours that appeared roughly were the illusory contours would be. To make the real condition comparable to the illusory condition, half of a real luminance-defined contour appeared at a time (as shown in Fig. 4C). A half real contour appeared exactly when the figure corner to which it's closest overlapped with a black oval. For example, the bottom half of the left real contour appeared exactly when a black oval overlapped with the lower left corner of the fat/thin rectangle; the top half of the same contour appeared exactly when the top left corner of the rectangle overlapped an oval. The same applied for the right contour. Because of the spacing of the ovals, the top and bottom half of either the left or right real contours never simultaneously appeared. Thus seeing complete contours in both the illusory and real conditions required accumulating information over time, but the illusory condition additionally involved contour interpolation.

2.1.3.3. Noise characteristics. The discriminated figures were embedded in static luminance noise following a Gaussian distribution. There was one noise field for all frames of a trial. The Gaussian noise distribution was truncated to ±2 standard deviations from the background luminance (as in Murray, Bennett, & Sekuler, 2005). The root-mean-square (RMS) contrast of a noise field (after truncation) was 13%. Noise "pixels" were composed of eight screen pixels (4 vertically × 2 horizontally) and had dimensions of 4 × 2 arcmin, creating a power spectral density of 46 mdeg².[5] Elongated dimensions of the noise pixels made it easier to capture the expected lower and higher spatial frequencies of the vertical and horizontal dimensions, respectively. Reducing the dimensionality of the input space in this way was expected to increase the signal-to-noise ratio (e.g., Ringach, Sapiro, & Shapley, 1997). To further increase SNR, we limited position uncertainty by decreasing the noise field area. The noise field

---

[5] A noise "pixel" will refer to a group of eight screen pixels that compose a unit of luminance noise, unless stated otherwise.
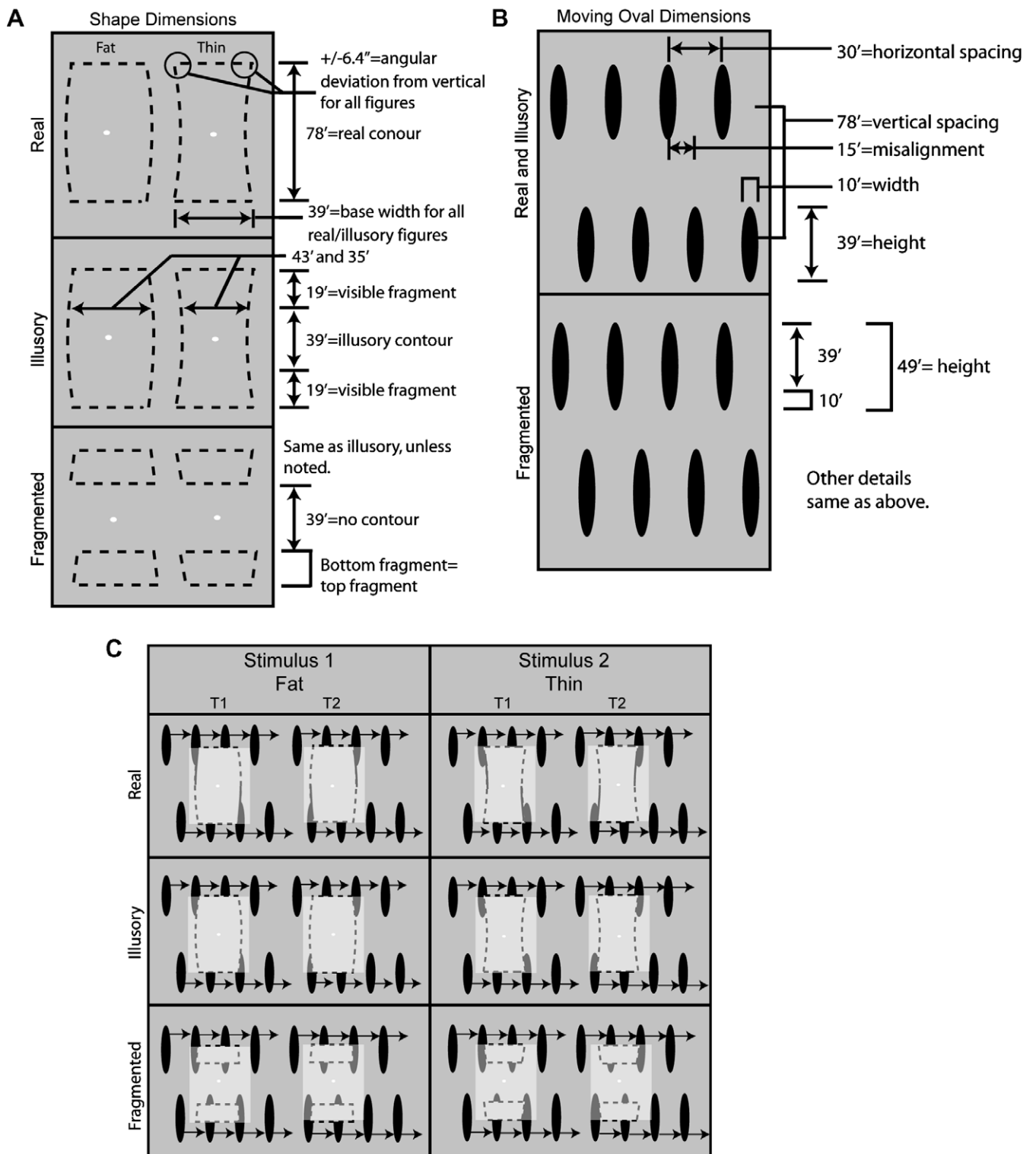
Fig. 4. Stimuli in Experiment 1. (A) Dimensions for fat/thin real, illusory and fragmented rectangles in Experiment 1. The dotted lines indicate the shape that subjects would typically see. (B) Dimensions of ovals that move behind the shapes. Ovals in the fragmented condition are elongated to produce a percept of disconnected fragments. Oval dimensions and spacing ensure that different parts of shapes become visible at different times. (C) A schematic representation of two frames (T1 and T2) from a fat and thin trial for each of the three conditions of Experiment 1. The translucent square represents the region that would be corrupted by static luminance noise. The arrows represent constant motion (8.4 deg/s) of the dark ovals.

measured 78 arcmin vertically by 60 arcmin horizontally so that the peripheral horizontal borders of the figures in all three conditions were coincident with the horizontal borders of the noise field (see translucent square regions in Fig. 4C). Consequently, the lower portion of the lower ovals and the upper portion of the upper ovals never were

affected by noise. In each trial, a noise field appeared exactly when the ovals moved.

2.1.4. Procedure

At the beginning of a trial, the black ovals emerged gradually from behind an otherwise invisible occluder. The ovals stopped moving exactly when all eight of the ovals no longer overlapped with the noise field. A white fixation point appeared in the center of the noise field and was present at all points during a trial. It was emphasized to subjects at the beginning of the experiment that it was "extremely important" to keep focused on the fixation point, even if they felt that that would degrade their performance. Subjects were also reminded of the importance of fixation after every 300 trials. The task in the illusory or real condition was to indicate via a button-press whether a fat or thin rectangle was presented. The task in the fragment condition was to indicate (via a button-press) whether the trapezoid fragments were fat or thin.

Signal contrast was modulated by altering oval darkness. A psychometric function determined a contrast level that would generate 70% performance. Specifically, we estimated threshold performance by running subjects on 40 trials per contrast level, with the following (Weber) contrast levels: 70%, 60%, 50%, 40%, 30%, and 20%. The selected contrast level remained fixed within a session of trials, but would be occasionally adjusted between sessions if performance deviated significantly from 70% (Murray et al., 2005, p. 142).

Trial types were blocked for the first two observers and each subject ran 8700 trials per block. For MC the order was: real, illusory, fragment. For observer BPK, the order was: illusory, fragment, real. Before each block, a subject received 40 high contrast practice trials of the condition. Four subjects/condition were subsequently added and each of these subjects began with 40 high contrast trials, and a psychometric function. Those subjects on average performed 2775 trials per condition on exactly one condition. The total number of trials per condition was 28,500, namely 8700 (BPK) + 8700 (MC) + 11100 (four additional naïve subjects). When subjects responded, a correct response was marked with a high beep and an incorrect response was marked with a low beep. To reduce fatigue, subjects received a 3-min, forced break after every 300 trials.

2.2. Dependent measures and data analysis

Two stimuli (S1, S2), denoting the "fat" and the "thin" figures, and two responses (R1, R2), denoting the "fat" and the "thin" responses, generated four possible stimulus–response categories for each trial, S1R1, S2R1, S1R2 and S2R2. To derive a raw classification image for a condition, we determined the average noise field for each of the four stimulus–response categories of that condition. Each of these noise fields was computed from trials across all subjects who ran that condition. These four average noise fields

were then combined: $CI = (S1R1 + S2R1) - (S1R2 + S2R2)$. On this coding schema, light CI regions denote a positive correlation between pixel contrast and a "fat" response; dark regions indicate a negative correlation. Next, each raw classification image was convolved with a 5×5 pixel kernel that is the outer product of $[1\ 1.6\ 3\ 1.6\ 1]^T$. The border regions that could not be accurately computed in the convolution span 8 screen pixels (9 arcmin) each on the top and bottom, and four screen pixels (4 arcmin) each on the left and right. Dotted white lines around the periphery in the CIs of Fig. 5 mark off these affected regions.

To examine the statistical significance of CI regions, convolved images were quantized with analytically derived thresholds (see Appendix A). In the quantized images, pixels significant at $p < .001$ are illustrated as black or white; pixels significant only at $p < .01$ are illustrated as faded black or off-white. All other pixels in the quantized image are deemed non-significant and are illustrated as mean gray (see Fig. 5).
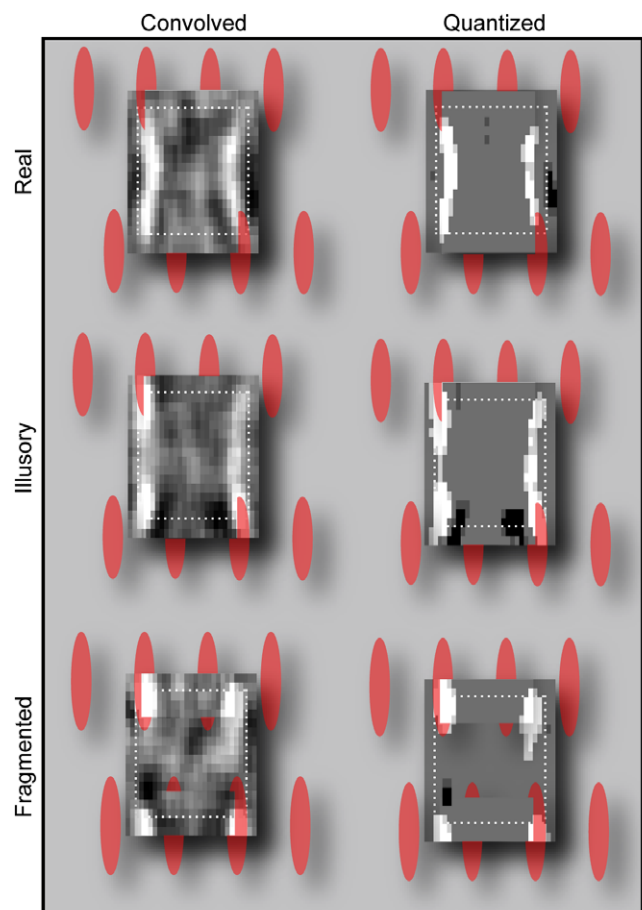
Fig. 5. Classification images from Experiment 1: Convolved and quantized CIs derived from six observers (five naive) from each of the three conditions in Experiment 1. As a signal landmark, superimposed ovals are shown, where one oval from each row is occluded by an unrotated fragment edge. The dotted line rectangle includes exactly the CI region not adversely affected by the convolution process. In the quantized images, pixels that are white or black are significant at $p < .001$; the off-white and faded-black pixels are significant only at $p < .01$.

2.3. Results and discussion

Weighted average performance levels (% correct) for all three conditions ranged between 72.3 (illusory) and 74.2 (real). The average signal (Weber) contrast was 32%, 54%, and 57% for the real, fragmented, and illusory condition, respectively. There was virtually no response bias; the percentage of "thin" responses for the real, illusory, and fragmented conditions was 51.2%, 51.0%, and 49.9%, respectively.

Classification images from Experiment 1 are shown in Fig. 5. In all conditions, noise regions strongly influenced response. Whereas only about five pixels would become significant by chance in a condition, the total number of significant pixels ($p < .01$) was 80 in the real condition, 140 in the illusory condition, and 82 in the fragmented condition. More importantly, regions near interpolated and luminance-defined boundaries affected performance comparably, with significant pixels appearing near discriminated boundaries in both cases. Regions between inducers were not active in the absence of real or interpolated contours. Consequently, even when inducers are presented gradually and sequentially, and so even when contour information must be accumulated over time, interpolation regions appear to influence performance. These results are very much in accord with those yielded with spatial displays (as shown in Fig. 3), and suggest an important commonality between spatial and spatiotemporal contour interpolation.

Another noteworthy result from this experiment is the fact that subjects could perform the task at all. In the illusory condition, subjects could effortlessly perceive and discriminate fat and thin high-contrast figures, even though the average total duration that at least one pixel appeared on an inducing edge was 71 ms. The ability to interpolate in these circumstances probably owes to spatiotemporal receptive fields that integrate over durations of 100 ms (Barlow, 1958; Burr, 1981; Burr, Ross, & Morrone, 1986; Ross & Hogben, 1975), and also to iconic memory mechanisms that subsequently store orientation (Von Wright, 1968) and shape information (Turvey & Kravetz, 1970). Such mechanisms may effectively prolong and stabilize the appearance of a physically transient stimulus, and thereby allow the visual system to interpolate with less information (Palmer et al., 2006, pp. 527–531).

3. Experiment 2: Spatiotemporal contour interpolation for moving objects

CIs from Experiment 1 showed that pixels near interpolated and real contours paths affected performance comparably, but those same pixels did not have an affect when contours were absent. These results show the functional consequences of spatiotemporal interpolation, and suggest a fundamental commonality with spatial interpolation, at least when figures are stationary. In Experiment 2, we examine if the same conclusions hold when figures move. Specifically, parts of a real, illusory, or fragmented figure were presented sequentially, as in Experiment 1, but figures translated across dark stationary background elements. Interpolation in this second experiment required that a subject both store and update the position of previously viewed fragments, so that they could be related to subsequently appearing fragments. To capture how this process unfolds, rather than having one noise field per trial and one CI per condition—as in Experiment 1, there was a new static noise field for each of the nine frames of movement (frame rate = 17 Hz), and a CI was computed for each of those frames (to produce a CI "movie"). Although dynamic noise dramatically increased the number of trials needed to derive reasonable CIs, it had the advantage of providing "snapshots" of how an object was represented from moment to moment. If subjects treat illusory contours like real contours and differently from fragmented contours, then a fundamental characteristic of spatiotemporal interpolation will have been uncovered. Such a finding would further motivate the conjecture that spatial contour interpolation is a special case of spatiotemporal contour interpolation.

3.1. Methods

3.1.1. Subjects

Seventy-five UCLA students who were naive to the purposes of the experiment participated for class credit in one condition for one or two sessions. Each observer ran between one and two hours and completed between 1400 and 3500 trials. The total number of trials completed across all naïve subjects was 45,000 per condition. In addition, the first author performed 15,000 trials for each of the three conditions. Thus, each condition involved exactly 26 subjects and 60,000 trials. All observers reported normal or corrected-to-normal vision.

3.1.2. Apparatus

The apparatus was the same as Experiment 1.

3.1.3. Stimuli

All trials involved six dark stationary rectangles—three on bottom, and three on top (see Fig. 6). Rectangles in a row were equally spaced from one another by 24 arcmin. To help create the sequential appearance of the figure edges, the left-most top rectangle was horizontally misaligned from the left-most bottom rectangle by 12 arcmin. In the illusory and real conditions, the rectangles measured 10 arcmin by 30 arcmin. In the fragmented condition, the lower portion of the top rectangles and the upper portion of the bottom rectangles were extended by 8 arcmin to create the appearance of distinct fragments.

Figures in all conditions were created by two vertically aligned gray trapezoids, the left sides of which incorporated right-angles and were invisible. In the illusory and

A Moving Shape Dimensions

Fat | Thin

Real
78′=real contour
+/-6.4° =angular deviation from vertical for all figures

Illusory
15′=visible fragment
48′=illusory contour
15′=visible fragment

Fragmented
Details same as illusory condition, unless noted.
48′=no contour
bottom visible fragments of illusory condition swapped.

B Static Rectangle Dimensions

Real and Illusory
24′=horizontal spacing
30′=height
10′=width
12′=misalignment
78′=vertical spacing

Fragmented
30′
8′
38′= height

All other details same as above.

C

Frame Number
1　2　3　4　5　6　7　8　9

Real — Fat
Real — Thin
Illusory — Fat
Illusory — Thin
Fragmented — Fat
Fragmented — Thin

Time Elapsed (ms)
0　59　118　176　235　294　353　412　471　529

real conditions, the central horizontal contour of a trapezoid was coincident with the central horizontal contours of a row of black rectangles. In those same conditions, each trapezoid pair either tapered inward toward the center or bulged slightly outward peripherally to provide the appearance of a single rectangle that bulged outward or tapered inward on the right side. In the fragmented condition, the stationary rectangles were elongated to produce the percept of two trapezoids, the right sides of which either sloped downward and to the right (labeled "fat") or sloped downwards and to the left ("thin").

In all trials, the top trapezoid measured 15 arcmin vertically, and the lower horizontal contour was 2 arcmin shorter than the upper horizontal contour for a thin response and was 2 arcmin longer for a fat response. Thus, the non-right angles of a trapezoid were ±6.4 deg from the vertical. In all trials, the distance between the horizontal, peripheral contours of the bottom and top trapezoid was 78 arcmin. In the illusory and real conditions, the bottom trapezoid was a reflection about the central horizontal axis of the top trapezoid. The fat and thin bottom trapezoids in the fragmented condition were the thin and fat bottom trapezoids of the illusory condition, respectively. This ensured that each frame of the two conditions would contain the same discriminated visible contour edges (across fat and thin trials). In all trials, the right vertices of the trapezoids expanded to the right at 2.1 deg/s, whereas the left side of the trapezoid remained stationary.

In the illusory and real conditions, the translation of the right vertices of each trapezoid yielded the impression of a contour of a single figure translating from left to right. In the fragmented condition, the impression was of two trapezoids the right vertices of which translated from left to right. The real condition differed from the illusory condition in that the real contour information appeared sequentially over time along with the inducers. The top half of the full vertical contour appeared during frames when the top inducer was visible, and the bottom half of the full contour appeared when the bottom inducer was visible, as shown in Fig. 6C. In all trials, there were nine frames (59 ms/frame) in a motion sequence, though the signal was always absent in the first and ninth frames. In all conditions, each task-relevant frame either provided shape information on the top half of the figure, or provided information on the bottom-half of the figure, but never provided both kinds of information. Although the displays were somewhat complicated, and although the fragments that produced interpolation were relatively sparse, naive observers were able to see

the contours quickly, with some subjects picking up the task on the first trial of practice.

In all trials, the noise field measured 78 by 58 arcmin and covered all but the outer 15 arcmin of the dark rectangles. The noise was dynamic in that there was a new static noise field for each of the nine frames of the motion sequence. Slower frame-rates were used, because otherwise CI SNR would be greatly reduced (Xing & Ahumada, 2002). At fast frame-rates, there might be significantly greater uncertainty as to what points match between frames, possibly because of the introduction of non-additive internal noise (Barlow & Tripathy, 1997; Lu & Liu, 2006; Morrone, Burr, & Vaina, 1995). Within a frame, the noise had a RMS contrast of 11% (after truncation), creating a power spectral density of 35 mdeg^2.

3.1.4. Procedure

The task in the illusory or real condition was to indicate via a button-press whether a fat or thin rectangle was presented. The task in the fragment condition was to indicate (via a button-press) whether the trapezoid fragments were fat (sloping downward and to the right) or thin (sloping downward and to the left). Task difficulty was modulated by altering the contrast of the dark rectangles. Every 400 trials a subject was required to take a 1-min break to avoid fatigue. Observer BPK performed (in chronological order) the illusory, fragmented, and real conditions. All other aspects of the procedure were the same as Experiment 1.

3.2. Dependent measures and data analysis

Individual data were combined, as described in Experiment 1. The same analyses were used as in Experiment 1, except that we calculated a classification image for each of the nine frames in the motion sequence.

3.3. Results and discussion

Weighted average performance levels (% correct) for all three conditions ranged between 73.3 (real) and 73.6 (illusory). There was virtually no response bias; the percentage of "thin" responses for the real, illusory, and fragmented conditions was 50.6%, 51.4%, and 49.7%, respectively. The average signal (Weber) contrast was 36%, 41%, and 48% for the real, fragmented, and illusory condition, respectively.

Classification images for Experiment 2 are shown in Fig. 7 and the numbers of significant pixels per frame are

Fig. 6. Stimuli from Experiment 2. (A) Dimensions for fat/thin real, illusory, and fragmented figures in Experiment 2. The moving figures are dotted because they are invisible unless they overlap with the dark background rectangles. (B) Dimensions of dark rectangles over which figures move. Background rectangles in the fragmented condition are elongated to produce a percept of disconnected fragments. In all conditions, the top and bottom part of a figure never simultaneously overlap with a background rectangle to produce contour information along a vertical edge. (C) A schematic representation of the two trial types at all nine frames of a trial for each of the three conditions. The duration of each frame was about 59 ms, and the time course is shown for reference. The translucent square represents the region that would be corrupted by dynamic luminance noise, and the arrows represent constant motion of the dotted figure.
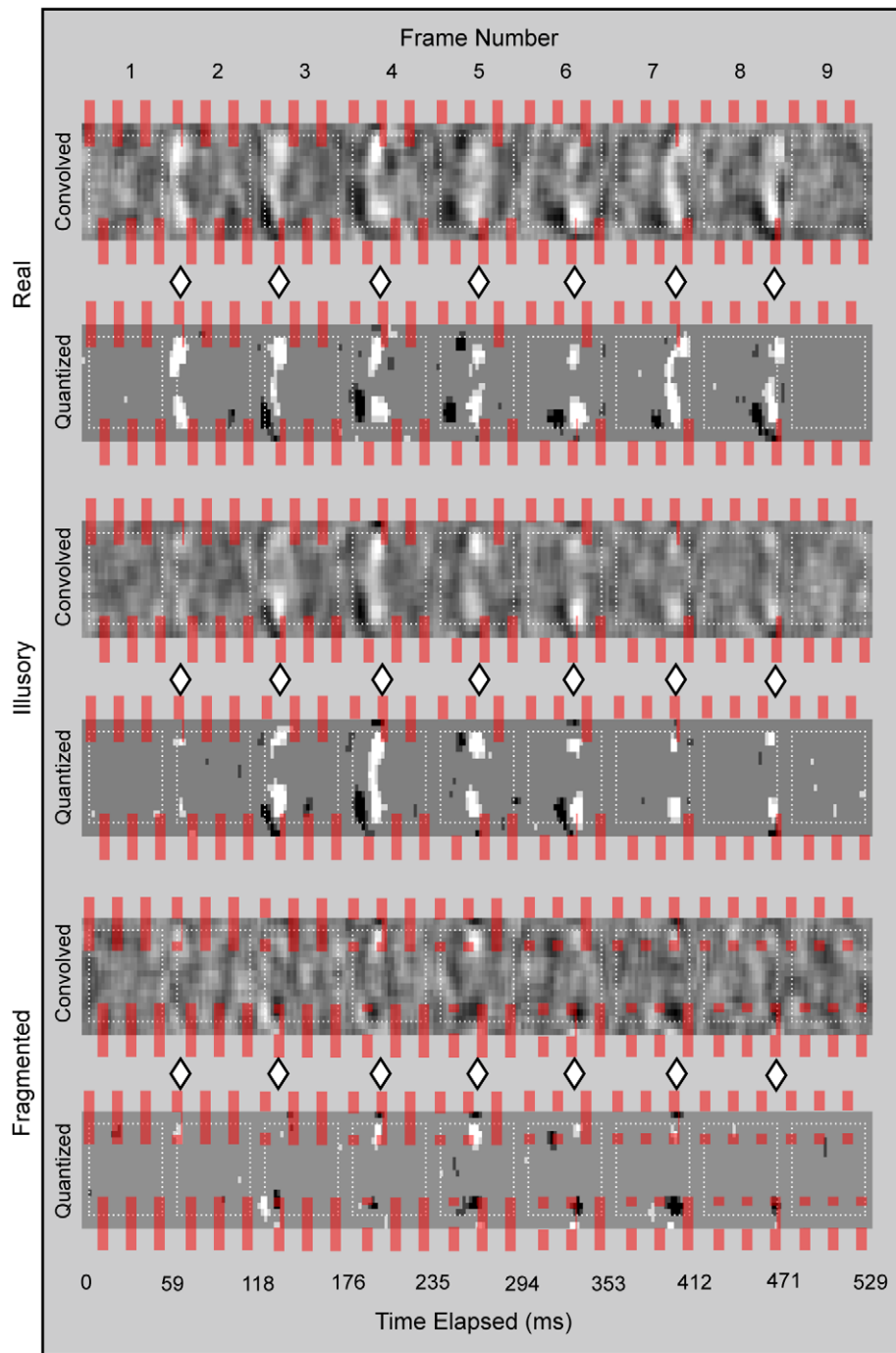
Fig. 7. Classification images from Experiment 2: Convolved and quantized classification images and time courses for each of the nine frames of motion are shown for each condition. The first and last frames never contained signal, and consequently do not contain many significant pixels. The superimposed red elements show the average position of a fragment against the dark background rectangles for that frame and condition. The dotted lines around the border of each frame mark the peripheral regions that could not be properly calculated in the convolution. The diamonds indicate the average position of inducer edges for a frame and condition. In the quantized images, pixels that are white or black are significant at $p < .001$; the off-white and faded-black pixels are significant only at $p < .01$.

shown in Fig. 8. Pixels in all three conditions influenced response at a rate much greater than chance, with the greatest number of pixels influencing a response in frame 4 in the illusory and real conditions, and in frame 5 in the fragmented condition. More importantly, in certain frames (especially frame 4), regions between inducers influenced performance in the illusory condition comparably to when there were luminance-defined contours. Those same regions were not active in the absence of real or interpolated boundaries. Consequently, when contours move and when inducers appear sequentially, interpolation can be observed to have functional effects similar to real contours.
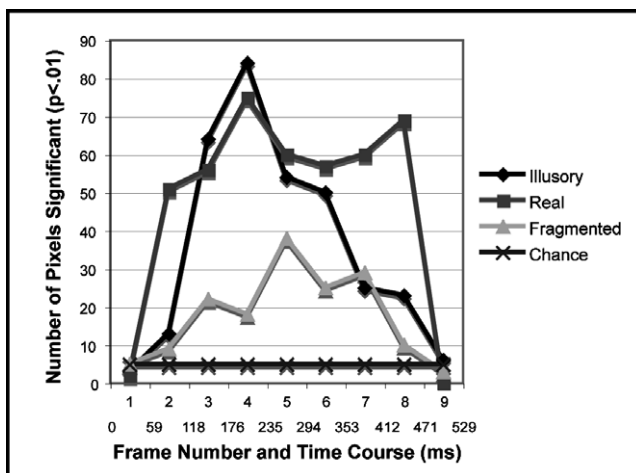
Fig. 8. Significant pixels in Experiment 2: Number of significant pixels as a function of frame number and time course (ms). Whereas only about five pixels would become significant by chance in each frame, many more pixels influenced response in frames of each condition.

These results resemble those found with spatial interpolation displays (Gold et al., 2000; see Fig. 3; Gold & Shubel, 2006) and, again, suggest a fundamental similarity between the two.

We note several other aspects of the CIs. As shown in Fig. 7, dominant CI features are positioned closely to the moving edge of an inducer, suggesting that responses were fairly tied to the signal. Also, pixels along the entirety of the illusory contour became most influential between 118 and 176 ms after the appearance of the first (top) inducer. This is consistent with the 175 ms time course estimate for spatial contour interpolation (Gold & Shubel, 2006). Finally, in the real condition a complete CI contour formed in frames 3 and 7, and in the fragmented condition, a light CI region near the top inducer and a dark CI region near the bottom inducer appeared together in frames 4 through 7. (The bottom portion of the fragmented CI has a polarity opposite to the illusory CI because the bottom fat and thin fragments of the two conditions were anti-correlated.) Since the top and bottom parts of a figure never became visible in any one frame in any condition, the hypothesized mechanisms of iconic storage and position updating (Palmer et al., 2006) appear not to be specific to spatiotemporal contour interpolation.

4. General discussion

In two experiments, subjects discriminated fat and thin noise-corrupted figures, the tops and bottoms of which were connected by illusory contours, luminance-defined contours, or no contours at all. In Experiment 1, the figures were stationary and became visible by gradually occluding and disoccluding dark moving background elements; in Experiment 2, the figures moved and became visible by gradually occluding and disoccluding dark stationary background elements. In both experiments, the distribution of the background elements ensured that the top and bottom parts of a discriminated contour never appeared together at once. Producing a complete figural percept in Experiment 1 required storing fragment information, while Experiment 2 additionally required updating the positions of fragments. CIs from both experiments showed that pixels along spatiotemporally interpolated boundaries influenced responses comparably to when real contour information appeared along those same boundaries. By contrast, in displays that did not support interpolation, there was little indication of pixel influences in regions between visible figure fragments. These results, taken together, suggest that spatiotemporally interpolated boundaries have consequences that go beyond a phenomenal presence.

In the following, we will discuss the relation between spatial and spatiotemporal contour interpolation, and the microgenesis of illusory contours as revealed in Experiment 2. We then consider an objection to using CIs as a method for revealing early or mid-level visual processing. We conclude with suggestions for future research.

4.1. Spatial contour interpolation as a limiting case of spatiotemporal contour interpolation

Because our CIs resemble those shown previously with spatial displays (i.e., Gold et al., 2000), an interesting possibility is that spatial interpolation may be a limiting case of spatiotemporal interpolation. Spatiotemporal interpolation is hypothesized to involve storing information about momentarily invisible fragment edges, and (if necessary) updating the positions of those edges for the purpose of interpolation with visible fragment edges. Here, we suggest that spatial interpolation occurs in cases where minimal information is stored, and where there is no position updating. There seems to be growing consensus that form and motion information are combined early. When subjects detected the drift direction of sinusoidal gratings masked by reverse phase gratings of various spatial and temporal frequencies, there was a coupling between the spatial and temporal frequency tuning functions, leading the authors to infer the operations of spatiotemporal receptive fields (Burr et al., 1986). More recently, when subjects identified objects moving behind a surface with narrow slits, spatial frequencies theoretically inaccessible from static views were utilized to decide global form and motion (Nishida, 2004). It was hypothesized that form and motion are entwined at very early stages, partly as a result of the direction and orientation specificity of certain V1 cells (Emerson, Bergen, & Adelson, 1992; Hubel & Wiesel, 1968). In a different study, when subjects determined whether partially occluded outline objects orbited clockwise or counterclockwise, performance depended on relatively low-level properties such as contrast, and temporal and spatial frequency (Lorenceau & Alais, 2001). These properties were inferred to influence the perception of global figure motion at early processing stages. Lennie (1998) argued on the basis of cortical

organization that form and motion (among other features) are coupled at early levels in processing. Although none of the foregoing studies specifically address spatiotemporal contour interpolation, they all appear to agree that determining the dynamic properties of an object is intimately associated with determining its form, and that this link occurs very early on in visual processing. Given that contour interpolation is itself a low-level process that is used for determining form (Grosof, Shapley, & Hawken, 1993; Halgren, Mendola, Chong, & Dale, 2003; Lee & Nguyen, 2001; Von der Heydt, Peterhans, & Baumgartner, 1984), it is reasonable to think that contour interpolation depends on spatiotemporal mechanisms that handle static arrays as a limiting case. Such a view fits with other recent data (e.g., Keane, Kellman, & Elwell, 2007; Palmer et al., 2006) and makes ecological sense, in that visual systems likely evolved to serve the needs of observers who perceive moving objects or who perceive stationary scenes during self-motion (Gibson, 1966, 1979).

4.2. Time course of contour interpolation

A number of paradigms have been used to estimate the time course of illusory contour formation. Gold and Shubel (2006) cross-correlated CIs from an illusory condition and a blurred ideal observer template from a real condition (with inducer regions removed). The correlation between the two images reached a peak value at about 175 ms and interpolation was inferred to require about the same amount of time to complete. Guttman and Kellman (2004) found that when observers determined whether a dot fell inside or outside a Kanizsa figure, precision and accuracy reached an asymptote at about 120–140 ms of stimulus presentation; contour interpolation was inferred to complete within the same period. Other psychophysical (e.g., Reynolds, 1981; Ringach & Shapley, 1996) and neuroimaging studies (Murray, Foxe, Javitt, & Foxe, 2004; for a review, see Seghier & Vuilleumier, 2006) have provided similar estimates for the completion of illusory contours. The CI results from Experiment 2 indicate a time course that is consistent with the foregoing. If we assume that (a) interpolation begins once the first inducer appears, (b) noise pixels become most influential mid-frame, and (c) pixels exercise influence when their locations are incorporated into a representation, then illusory contours appear to form within 147 ms (since a CI contour first appears in the fourth CI frame).

Relative differences between CI conditions also fit with previous data on time course. Pixels became influential earlier in the real than in the illusory condition (see also, Fig. 8) and this corroborates previous findings that real contours are processed more quickly than their illusory contour counterparts (Gold & Shubel, 2006; Guttman & Kellman, 2004). Moreover, in the fragmented condition, CI pixels near both inducers were not clearly used until frame 4—a frame later than the illusory condition. This supports previous findings that information is extracted

more quickly from relatable than from unrelatable object parts (Moore, Yantis, & Vaughan, 1998).

It should be pointed out in passing that although others have inferred absolute time course from CIs (Gold & Shubel, 2006), and although our estimate of 147 ms fits with previous estimates, the underlying assumption—that pixels begin biasing a response when a representation first forms—may not be true. For example, at t_1 a pixel luminance value may appear at a point where a contour has not yet formed, and by t_2 this pixel information may be extrapolated (along with inducer information) to another point over which a contour *has* formed (Changizi & Widders, 2002; Nijhawan, 1994; Palmer et al., 2006). In this case, a CI pixel would show up as significant at one location in one frame, even though it influenced a representation that formed at a different (updated) location in the next frame. Inferring time course from CIs as we have done could also conceivably lead to an *under*estimation of processing speed. Pixels appearing within one frame (f_2) could bias performance by altering (perhaps unstable) representations initially formed in the previous frame (f_1), and this bias could reduce or overshadow the pixel influences found in (f_1). Visual *post-diction* is evidenced in backward masking (Bachmann, 1994), and position mislocation (Eagleman & Sejnowski, 2000), and may also occur during spatiotemporal figure formation. Although we are not in a position to rule out extrapolation or post-diction, there is little reason to suppose that either process is more prevalent in one condition than another. Consequently, we regard classification imaging as a useful tool for uncovering relative, if not absolute, time courses for object formation.

4.3. Do fat/thin CIs reveal contour interpolation?

Gold et al. (2000) and Sekuler and Murray (2001) maintained that fat/thin CIs reveal "behavioral receptive fields" that correspond to perceptually completed contours. However, some have questioned whether CIs truly reveal low-level filling-in processes. Gosselin and Schyns (2003) showed that when subjects were asked to "detect" a non-existent "S" in white noise over the course of 20,000 trials, and when a CI was calculated (with only two stimulus–response categories), the resulting images displayed the target letter. Might fat/thin CI features also derive from cognitively driven top-down influences? We address this concern by first acknowledging that attention and cognitive strategy do affect low-level processing. Just as modulating attention can affect the perception of motion coherence (Liu, Fuller, & Carrasco, 2006), contrast (Pestilli & Carrasco, 2005), and lightness (Tse, 2005), so too will it affect contour interpolation (e.g., Montaser-Kouhsari & Rajimehr, 2004). Nevertheless, there are a number of reasons to believe that fat/thin CIs clearly show contour interpolation effects. First, the fat/thin paradigm itself has repeatedly yielded results that are best explained in terms of interpolation (Kellman, Yin, & Shipley, 1998; Ringach & Shapley, 1996). Displays consisting of relatable fragments can be

discriminated more quickly and more accurately than a number of non-relatable controls, and only relatable displays are affected by irrelevant lines placed between discriminated edges. These data cohere with those of Gold et al. (2000) and suggest that—at least in spatial displays—irrelevant information appearing near interpolation paths automatically affect how subjects interpolate shape. Second, in our experiments, subjects received feedback. Rather than using the feedback to form approximately ideal templates, subjects continued to rely upon the information-less interpolation regions, indicating, again, that the CI results are primarily stimulus-driven. Third, averaging across subjects should result in incoherent CIs when subjects are not sufficiently constrained by the stimuli (Gosselin & Schyns, 2003, pp. 506–508); in our case, averaging across subjects resulted in more salient CI features in all three conditions. Fourth, in Experiment 2, interpolation effects appeared only at some temporal intervals. This would not obviously be expected if a cognitive template maintained throughout the presentation sequence determined the observer's responses. The fact that interpolation estimates furnished by the CI data approximate earlier estimates of interpolation or the fact that CI contour regions become active earlier in the real condition suggests that CIs are tapping into early processing (Guttman & Kellman, 2004). Finally, other CI researchers tentatively endorse classification imaging for understanding low-level processes. Neri and Levi (2006), for example, claim that the evidence so far suggests that "noise classification appears to target stages reflecting computations that are very similar to those observed in physiological recordings", although they think that more research is necessary before firm conclusions can be made (pp. 2470–2472).

4.4. Future directions

The experiments presented in the present paper suggest a number of avenues for future research, two of which will be outlined here. First, the cases we examined consider how the appearance of noise near interpolated boundaries affects interpolated shape, but the opposite causal relation is equally interesting. If subjects have to identify whether a small target is lighter or darker than the background, responses may be slower or less accurate when the targets appear near interpolated boundaries of fat or thin figures. Filling-in processes in apparent motion affect how observers respond to a target (Yantis & Nakama, 1998), and a similar effect might occur for interpolated contours.

Another extension to the present work would be to examine the effects of noisy interpolation regions on the perception of spatiotemporally presented occluded (amodal) figures. CIs show that both modally and amodally completed figures are corrupted by noisy interpolation regions in spatial displays (Gold et al., 2000), and so it is reasonable to expect similar results to obtain in spatiotemporal displays. Such results would lend further support to

the claim that modal and amodal completion are subserved by common interpolation mechanisms (Kellman, Garrigan, & Shipley, 2005, 2007; Kellman et al., 1998; Shipley & Kellman, 1992; though see Anderson, 2007). Exploring the relationship between boundary formation contexts—whether modal and amodal or spatial and spatiotemporal—will ultimately aid in categorizing and understanding the processes central to forming representations of coherent and persisting objects.

Appendix A

The standard deviation (σ) of a filtered CI can be computed analytically as follows. According to the derivation by Murray, Bennett, and Sekuler, 2002 (details on p. 83), when the CI is calculated by combining the four average noise fields as (S1R1 + S2R1) − (S1R2 + S2R2), the variance of the CI is $\sigma_c^2 = (\frac{1}{n_{11}} + \frac{1}{n_{21}} + \frac{1}{n_{12}} + \frac{1}{n_{22}})\sigma_N^2$, in which n_{SR} denotes the number of trials in each stimulus–response category, and σ_N^2 is the variance of external noise (after truncation). The filtered CI is computed by convolving the kernel, $\mathbf{K}$, with the original C, $\mathbf{C}$, as $\mathbf{K} * \mathbf{C}$. Furthermore, the variance of the filtered CI can be derived as $\sigma^2 = \mathrm{VAR}(\mathbf{K} * \mathbf{C}) = \sum_{i=1}^{w}\sum_{j=1}^{w}k(i,j)^2\sigma_c^2 = \sum_{i=1}^{w}\sum_{j=1}^{w}k(i,j)^2(\frac{1}{n_{11}} + \frac{1}{n_{21}} + \frac{1}{n_{12}} + \frac{1}{n_{22}})\sigma_N^2$, where $\mathbf{w}$ is the kernel size. Accordingly, the variance of the filtered CI is determined by the kernel weights, the number of trials in each stimulus–response category and the variance of added noise. The foregoing analytical determination of the standard deviation was confirmed via a bootstrapping method.

References

Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision, 2*(1):5, 66–78. Available from: http://journalofvision.org/2/1/5/, doi:10.1167/2.1.5.

Ahumada, A. J. (1996). Perceptual classification images from Vernier acuity masked by noise [Abstract]. *Perception, 26*(Suppl. 18), 18.

Ahumada, A. J., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America, 49*, 1751–1756.

Anderson, B. L. (2007). The demise of the identity hypothesis and the insufficiency and nonnecessity of contour relatability in predicting object interpolation: Comment on Kellman, Garrigan, and Shipley (2005). *Psychological Review, 114*(2), 470–487.

Bachmann, T. (1994). *Psychophysiology of visual masking*. Commack, NY: Nova Science.

Barenholtz, E., & Feldman, J. (2006). Determination of visual figure and ground in dynamically deforming shapes. *Cognition, 101*(3), 530–544.

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *Journal of Physiology, London, 141*, 337–350.

Barlow, H., & Tripathy, S. P. (1997). Correspondence noise and signal pooling in the detection of coherent visual motion. *Journal of Neuroscience, 17*(20), 7954–7966.

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

Bruno, N., & Bertamini, M. (1990). Identifying contours from occlusion events. *Perception & Psychophysics, 48*, 331–342.

Bruno, N., & Gerbino, W. (1991). Illusory figures based on kinematics. *Perception, 20*(2), 259–273.

Burr, D. C. (1979). *On seeing objects in motion*. Doctoral dissertation. University of Cambridge.

Burr, D. C. (1981). Temporal summation of moving images by the human visual system. *Proceedings of the Royal Society of London B, 211*, 321–339.

Burr, D. C., & Ross, J. (1986). Visual processing of motion. *Trends in Neuroscience, 9*, 304–306.

Burr, D. C., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London B, 227*, 249–265.

Changizi, M. A., & Widders, D. M. (2002). Latency correction explains the classical geometrical illusions. *Perception, 31*(10), 1241–1262.

Eagleman, D. M., & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science, 287*(5460), 2036–2038.

Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research, 32*, 203–218.

Fantoni, C., & Gerbino, W. (2003). Contour interpolation by vector-field combination. *Journal of Vision, 3*(4):4, 281–303. Available from: http://journalofvision.org/3/4/4/, doi:10.1167/3.4.4.

Gibson, J. J. (1966). The problem of temporal order in stimulation and perception. *Journal of Psychology, 62*, 141–149.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton.

Gold, J. M., Murray, R. M., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology, 10*, 663–666.

Gold, J. M., & Shubel, E. (2006). The spatiotemporal properties of visual completion measured by response classification. *Journal of Vision, 6*(4):5, 356–365. Available from: http://journalofvision.org/6/4/5/, doi:10.1167/6.4.5.

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science, 14*, 505–509.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Grosof, D., Shapley, R., & Hawken, M. (1993). Macaque V1 neurons can signal "illusory" contours. *Nature, 365*, 550–552.

Guttman, S. E., & Kellman, P. J. (2004). Contour interpolation revealed by a dot localization paradigm. *Vision Research, 44*, 1799–1815.

Halgren, E., Mendola, J., Chong, C. D., & Dale, A. M. (2003). Cortical activation to illusory shapes as measured with magnetoencephalography. *Neuroimage, 18*(4), 1001–1009.

Helmholtz, H. v. (1867/1962). *Treatise on physiological optics* (Vol. 3). New York: Dover Publications.

Hubel, D. H., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215–243.

Kandil, F. I., & Lappe, M. (2007) Spatio-temporal interpolation is accomplished by binocular form and motion mechanisms. PLoS ONE, 2(2): e264. doi:10.1371/journal.pone.0000264.

Keane, B. P., Kellman, P. J., & Elwell, C. M. (2007). Classification images reveal differences between spatial and spatiotemporal contour interpolation [Abstract]. J*ournal of Vision, 7*(9), 603, 603a. Available from: http://journalofvision.org/7/9/603/, doi:10.1167/7.9.603.

Kellman, P. J., & Cohen, M. H. (1984). Kinetic subjective contours. *Perception & Psychophysics, 35*, 237–244.

Kellman, P. J., Garrigan, P., & Shipley, T. F. (2005). Object interpolation in three dimensions. *Psychological Review, 112*(3), 586–609.

Kellman, P. J., Garrigan, P., Shipley, T. F., & Keane, B. P. (2007). Interpolation processes in object perception: Reply to Anderson (2007). *Psychological Review, 114*(2), 488–508.

Kellman, P. J., Guttman, S. E., & Wickens, T. D. (2001). Geometric and neural models of object perception. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 183–245). Amsterdam: Elsevier.

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

Kellman, P. J., Yin, C., & Shipley, T. F. (1998). A common mechanism for illusory and occluded object completion. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 859–869.

Lee, S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 1907–1911.

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27*, 889–935.

Liu, T., Fuller, S., & Carrasco, M. (2006). Attention alters the appearance of motion coherence. *Psychonomic Bulletin & Review, 13*, 1091–1096.

Lorenceau, J., & Alais, D. (2001). Form constraints on motion binding. *Nature Neuroscience, 4*(7), 745–751.

Lu, H., & Liu, Z. (2006). Computing dynamic classification images from correlation maps. *Journal of Vision, 6*(4):12, 475–483. Available from: http://journalofvision.org/6/4/12/, doi:10.1167/6.4.12.

Montaser-Kouhsari, L., & Rajimehr, R. (2004). Attentional modulation of adaptation to illusory lines. *Journal of Vision, 4*(6), 3, 434–444. Available from: http://journalofvision.org/4/6/3/, doi:10.1167/4.6.3.

Moore, C. M., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science, 9*, 104–110.

Morgan, M. J., Findlay, J. M., & Watt, R. J. (1982). Aperture viewing: A review and a synthesis. *Quarterly Journal of Experimental Psychology A, 34*, 211–233.

Morrone, M. C., Burr, D. C., & Vaina, L. M. (1995). Two stages of visual processing for radial and circular motion. *Nature, 376*, 507–509.

Murray, M. M., Foxe, D. M., Javitt, D. C., & Foxe, J. J. (2004). Setting boundaries: Brain dynamics of modal and amodal illusory shape completion in humans. *Journal of Neuroscience, 24*, 6898–6903.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision, 2*(1):6, 79–104. Available from: http://journalofvision.org/2/1/6/, doi:10.1167/2.1.6.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2005). Classification images predict absolute efficiency. *Journal of Vision, 5*(2):5, 139–149. Available from: http://journalofvision.org/5/2/5/, doi:10.1167/5.2.5.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Neri, P., & Levi, D. M. (2006). Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision Research, 46*, 2465–2474.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature, 370*, 256–257.

Nishida, S. (2004). Motion based analysis of spatial patterns by the human visual system. *Current Biology, 14*, 830–838.

Palmer, E. M. (2003). *Spatiotemporal relatability in the perception of dynamically occluded objects*. Doctoral dissertation, University of California, Los Angeles.

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.

Parks, T. E. (1965). Post-retinal visual storage. *American Journal of Psychology, 78*, 145–147.

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

Pestilli, F., & Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research, 45*, 1867–1875.

Reynolds, R. I. (1981). Perception of an illusory contour as a function of processing time. *Perception, 10*, 107–115.

Ringach, D. L., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research, 36*, 3037–3050.

Ringach, D. L., Sapiro, G., & Shapley, R. (1997). A subspace reverse-correlation technique for the study of visual neurons. *Vision Research, 37*, 2455–2464.

Ross, J., & Hogben, J. H. (1975). The Pulfrich effect and short-term memory in stereopsis. *Vision Research, 15*, 1289–1290.

Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception, 30*, 339–366.

Sekuler, A. B., & Murray, R. F. (2001). Visual completion: A case study in grouping. In T. F. Shipley & P. J. Kellman (Eds.), *From fragments to objects: Segmentation and grouping in vision* (pp. 265–294). New York: Elsevier.

Seghier, M., & Vuilleumier, P. (2006). Functional neuroimaging findings on the human perception of illusory contours. *Neuroscience Biobehavioral Reviews, 30*, 595–612.

Shipley, T. F., & Kellman, P. J. (1990). The role of discontinuities in the perception of subjective figures. *Perception and Psychophysics, 48*(3), 259–270.

Shipley, T. F., & Kellman, P. J. (1992). Perception of partly occluded objects and illusory figures: Evidence for an identity hypothesis. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 106–120.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*(11, Whole No. 498).

Tse, P. U. (2005). Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Research, 45*(9), 1095–1098.

Tse, P. U., & Albert, M. K. (1998). Amodal completion in the absence of image tangent discontinuities. *Perception, 27*(4), 455–464.

Turvey, M., & Kravetz, S. (1970). Retrieval from iconic memory with shape as the selection criterion. *Perception & Psychophysics, 8*, 171–172.

Von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). *Illusory contours and cortical neuron responses*. Science.

Von Wright, J. (1968). Selection in visual immediate memory. *Quarterly Journal of Experimental Psychology, 20*, 62–68.

Watamaniuk, S. N. J., & McKee, S. P. (1995). 'Seeing' motion behind occluders. *Nature, 377*, 729–730.

Xing, J., & Ahumada, A. J. (2002). Estimation of human-observer templates in temporal-varying noise [Abstract]. *Journal of Vision, 2*(7):343, 343a. Avaiable from: http://journalofvision.org/2/7/343/, doi:10.1167/2.7.343.

Yantis, S., & Nakama, T. (1998). Visual interactions in the path of apparent motion. *Nature Neuroscience, 1*, 508–512.