

# Rapid detection of salient regions: Evidence from apparent motion

Damian A. Stanley

Center for Neural Science, New York University, New York, NY, USA



Nava Rubin

Center for Neural Science, New York University, New York, NY, USA



Most studies that have used Kanizsa-type illusory figures to investigate perceptual completion have treated the crisp bounding illusory contours (ICs) and the enclosed region as nondissociable stimulus attributes. However, there is evidence that enclosed “salient regions” (SRs; Stanley & Rubin, 2003) are detected even in cases when bounding ICs are not perceptually completed. Here we used apparent motion (AM) to test whether SRs are detected in the absence of crisp bounding ICs. Kanizsa-type stimuli were modified in ways that eliminated the bounding ICs, but the clear impression of an enclosed region remained. SR stimuli were embedded in an array of like inducers. On successive frames, the inducers in the array rotated in a way that resulted in translation of the enclosed region. Four speeds of translation were tested. Observers performed a two-alternative forced-choice task on the direction of translation. Perceptually completed SRs produced robust AM whether they were bound by crisp ICs or not—observer performance was as good and, in certain cases, even better for SRs with no bounding ICs. We interpret these findings within a theoretical framework that makes a distinction between region-based and contour-based segmentation processes that operate in concert to achieve segmentation of the visual scene.

Keywords: illusory contours, salient regions, perceptual completion, segmentation, apparent motion

## Introduction

As we navigate through the world, our visual system must rapidly segment the retinal image into regions that correspond to distinct surfaces. Scene segmentation is a nontrivial task because surfaces that are whole in the real world are often fragmented in the retinal image because of occlusion, shadows, or unfavorable lighting conditions. “Illusory figures”—completed surfaces for which portions of the bounding contours are perceived in the absence of any luminance gradient (Kanizsa, 1955; see Figure 1a for an example)—have been used extensively to study the processes that give rise to perceptual completion of fragmented surfaces (for reviews, see Nieder, 2002; Spillmann & Dresch, 1995). Most of the studies have focused on the crisp bounding contours of illusory figures, paying less attention to the region that they enclose. This was perhaps because of a tacit assumption that the bounding illusory contours (ICs) and the enclosed region are not dissociable. However, the modified Kanizsa stimulus in Figure 1b illustrates that it is possible to eliminate the crisp bounding ICs while maintaining the impression of an enclosed region (Shipley & Kellman, 1990; Stanley & Rubin, 2003). The perceptual construct of a global region in Figure 1b occurs although it is clearly apparent that, in this case, it is an accidental consequence of local feature arrangement rather than an actual global surface. Is this merely a perceptual “error,” or could it have some function? Below we summarize a growing body of evidence for the continual

operation of processes that parse the image into major global regions, which we term *salient regions* (SRs). Furthermore, these *region-based* processes are functionally dissociable from *contour-based* processes responsible for the precise delineation of surface boundaries. The perception of SRs that do not correspond to contour-bound surfaces (as in Figure 1b) is one manifestation of this dissociation. Recent results from behavioral studies, functional magnetic resonance imaging (fMRI), and computer vision provide evidence for this dissociation in human vision and give it computational rationale.

Gurnsey, Poirier, and Gascon (1996) provided psychophysical evidence for the perceptual validity of SRs. The authors followed up on a study by Davis and Driver (1994), who used a visual search paradigm to demonstrate parallel detection of Kanizsa-type illusory figures embedded in an array of distractors (inducers rotated outward). Davis and Driver attributed this “pop-out” effect to early, automatic completion of the illusory figures, but did not distinguish between the bounding ICs and the perceptually completed surface. Gurnsey et al. (1996) made this distinction and showed that the “pop-out” effect remained even when they altered the stimuli to interfere with completion of the bounding ICs. They concluded that illusory contour completion was not responsible for the pop-out effect and suggested that the presence of the enclosed salient region (what they termed the “subjective figure”) might be.

More recently, Stanley and Rubin (2003) conducted an fMRI study that followed up on findings of Hirsch et al. (1995) and Mendola, Dale, Fischl, Liu, and Tootell (1999).

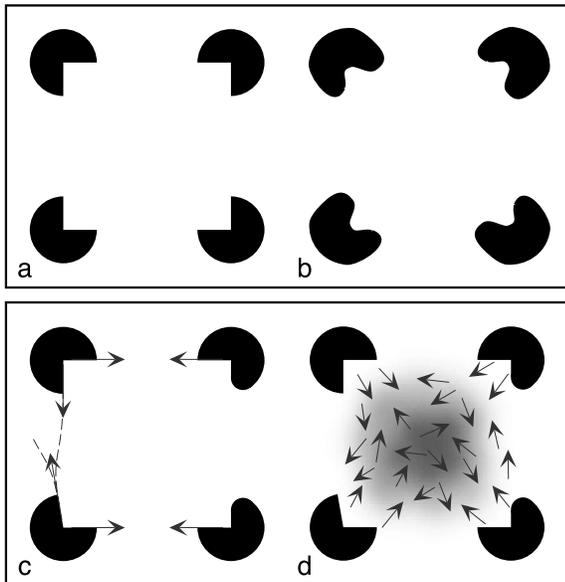


Figure 1. Contour-based and region-based completion. (a). Illusory Kanizsa square (Kanizsa, 1955): The perceptually completed (“illusory”) square is seen as bounded by a contour all around. (b). Salient region stimulus (SR): Slight modifications to the “pac-man”-shaped inducers eliminate the crisp bounding illusory contours (ICs), but the impression of an enclosed region is maintained. (c). Contour-based completion: Signals propagate along edges to complete portions of bounding contours which are missing in the image because of occlusion or lighting conditions (top & bottom arrows). Contour completion processes are dependent upon precise delineation (left side, weak IC) and junction structure (right side, no IC). (d). Region-based completion: Region-based processes seek to identify contiguous image regions that likely correspond to figural surfaces in the scene. Those processes propagate signals, that is, between a pixel (or a cluster of pixels) and all its neighbors, and evaluate similarity in surface properties (e.g., luminance, texture) for every neighboring pair. A salient region is one for which the similarity within region members is large, whereas the similarity to other neighboring pixels is small (for mathematical formulations see Sharon et al., 2000; Shi & Malik, 2000).

Those earlier studies showed that the lateral occipital complex (LOC)—a brain region previously shown to be involved in object perception (Malach et al., 1995; for reviews, see Grill-Spector, 2003; Grill-Spector, Kourtzi, & Kanwisher, 2001)—responded to the presence of Kanizsa-type illusory figures more than to controls in which the inducers faced outward. Stanley and Rubín (2003), in an analogous paradigm, showed that LOC responses to the SR stimulus in Figure 1b were similar in magnitude to LOC responses to Kanizsa-type illusory surfaces (Figure 1a). Thus, the LOC responds to SRs whether or not the regions are bounded by crisp illusory contours.

What advantage would the visual system gain from detecting salient regions in an image? A plausible explanation is offered by considering the computational demands of segmentation on the one hand and the ecology

of real-world scenes on the other. Determining the exact boundaries of an image region, its occlusion relationship to neighboring regions, and, ultimately, its status as a figural surface versus background region requires detailed processing of the contours and junction structure in the image (e.g., Grossberg & Mingolla, 1985; Guzman, 1969; Heitger, Rosenthaler, von der Heydt, Peterhans, & Kubler, 1992; Kellman & Shipley, 1991; Rubín, 2001; Shipley & Kellman, 1990; cf. Figure 1c). Performing such processing over the entire image can lead to daunting computational costs, as has been observed often in the computer vision literature (e.g., Mumford, 1994). This has led to the development of algorithms that perform a crude-but-fast parsing of the image and that detect regions that *likely* correspond to major objects in the scene—the computer vision definition of saliency (for mathematical definitions, see, e.g., Sharon, Brandt, & Basri, 2000; Shi & Malik, 2000; cf. Figure 1d). Although occasionally a region deemed salient may turn out to be a false alarm, actually being part of the background (e.g., as in Figure 1b), more often such regions do correspond to the main objects in the scene. Thus, it can be a useful strategy to direct the more computationally intensive contour and junction processing to a manageable number of regions in the image. Furthermore, region-based processes can be used to find figural surfaces when portions of the bounding contours are missing in the original image (e.g., for illusory surfaces) or when bounding contours are degraded because of poor visibility (e.g., blur, noise). Because a contiguous region is always bound by a closed contour (by necessity, topologically), the boundaries of the detected regions of highest saliency give a first approximation for the contour map of the image. This map can be subsequently crossed with edge information and further refined via a concerted operation of region-based and contour-based processes.

The study presented here provides further evidence for the perceptual validity of SRs and offers a method to quantify them—to measure “how salient” an image region is. The method is based on constructing image sequences in which (putative) SRs can serve as visual cues for motion correspondence. Ramachandran (1985, 1986) observed that illusory figures could exhibit apparent motion (AM) in a manner similar to “real” figures (ie, figures whose bounding contours are defined by luminance gradients). Specifically, Ramachandran (1985, 1986) showed that alternation between the two frames in Figure 2a results in the perception of an illusory square hopping back and forth and is not just local motion of the inducing elements. Bravo, Blake, and Morrison (1988) showed that cats were able to perform a two-alternative forced-choice task that relied on the perception of a hopping illusory square, lending further support to the assertions that (a) perceptual surface completion is a basic and robust process in the mammalian visual system and (b) perceptually completed figures can be treated by the visual system in the same way as luminance-defined figures (e.g., undergo correspondence for AM). This, in turn, suggests that AM displays may be

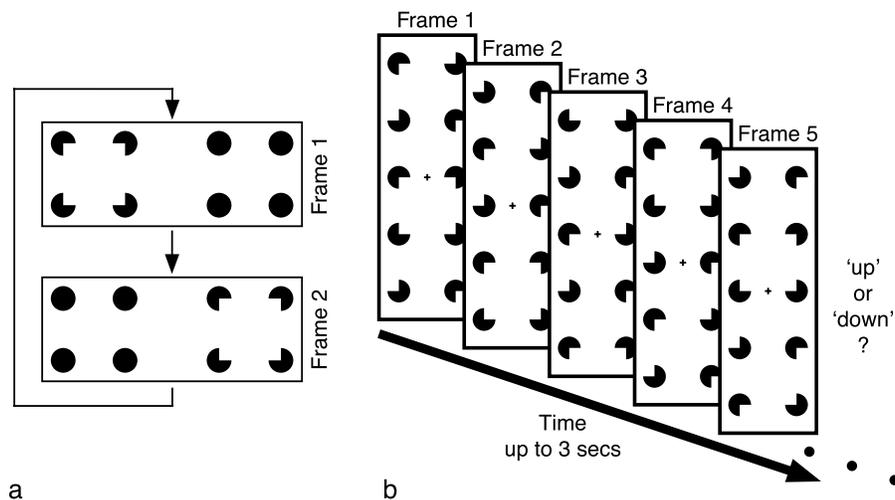


Figure 2. Apparent motion (AM) of perceptually completed surfaces. (a). The Kanizsa square is seen to move (“hop”) in front of two sets of filled disks when observers are presented with alternations between frames 1 and 2 (Ramachandran, 1985). (b). An example of the AM displays used in the experiments presented here. Observers performed a two-alternative forced-choice task on the direction of motion (up or down). The enclosed regions were one of eight stimulus types (cf. Figure 3; shown here is stimulus type 1, a Kanizsa square).

used to probe when a figure is perceptually completed. In the current study, we used AM of Kanizsa-type stimuli to probe for the detection of salient regions in the absence of crisp bounding illusory contours.

*A note about terminology:* In previous studies, authors often used the term “illusory contours” to refer to both the crisp bounding contours and to the region enclosed within those contours (e.g., the Kanizsa square). To adhere to the distinction we are making here between these two aspects of illusory figures, we reserve the term “illusory contours” (ICs) to refer solely to the crisp bounding contours of perceptually completed regions (when they exist); we use the unqualified term “salient region” (SR) to refer to an enclosed global region, whether or not it is bounded by ICs. In the few times when we need to distinguish whether an SR is bound by ICs or not, we will refer to it as *IC-bound* or *unbound*.

## Methods

### Observers

Six experienced psychophysical observers (3 men and 3 women, 6 right-handed, 23–31 years old) participated in the study. All had normal or corrected-to-normal vision. One observer was an author.

### Experimental procedure

We presented observers with displays in which an enclosed region was displaced between successive frames, either upward or downward. Observers indicated the

perceived direction of displacement (up or down) with a button press in a 2-alternative forced-choice task. Figure 2b shows an example of the stimulus sequence for the case of an IC-bound SR stimulus (a Kanizsa square). On each trial, five stimulus frames looped continuously until the observer responded or until the maximum viewing time of 3 s was reached. Trials were separated by a 2 s inter trial interval.

### Stimuli

In all, we tested seven different SR stimulus types (Figure 3a) that fell into 2 categories: SRs with bounding ICs (stimulus types 1 and 2) and SRs in which the bounding ICs had been disrupted in various ways (stimulus types 3–7; see caption of Figure 3a for detailed descriptions). The support ratio—the ratio between the luminance-defined portion and the entire bounding contour—for stimulus types 1 and 2 (IC-bound SRs) was 0.4 and 0.25, respectively. For the unbound SR stimulus types (3–7), the support ratio is difficult to quantify, because with no bounding contour, the shape and perimeter of the enclosed region are vague. Thus, we use the term *support ratio* only for stimulus types in which there is a putative enclosed square (e.g., when the inducers are aligned; stimulus types 3 and 6) to refer to the luminance-defined portion of the enclosed region that is directly tangent to the enclosed square. In addition to the seven SR stimulus types, we tested a control stimulus with a facing-out, symmetric inducer arrangement so that no enclosed region was present (stimulus type 8).

The display for each frame consisted of an array of 10 (5 vertical  $\times$  2 horizontal) inducers, 4 of which were arranged

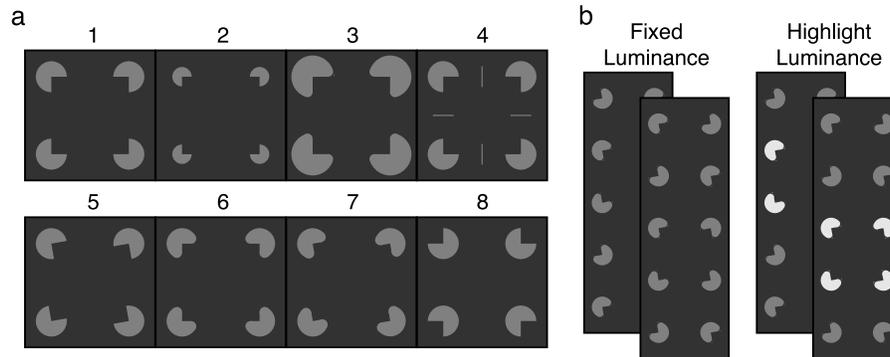


Figure 3. (a). The eight stimulus types used in the experiment: (1) Kanizsa square with a support ratio of 0.4. (2) Kanizsa square with a support ratio of 0.25. (3) Illusory contour completion is disrupted by rounding the L-junctions at the outer tips of the inducers (Shipley & Kellman, 1990). The support ratio is 0.4. (4) The line segments added orthogonally to the illusory portions of the bounding contour disrupt the perception of illusory contours (Gurnsey et al., 1996). (5) Illusory contour completion is disrupted by misaligning the inducers (Kellman & Shipley, 1991). (6) The same manipulation as in stimulus type 3, but with a lower support ratio of 0.25 (same as stimulus type 2). The resulting diameter of the inducers is the same as in stimulus type 1. (7) Similar to stimulus type 6, but IC completion is further disrupted by misaligning inducers. (8) No-SR control stimulus (note that symmetry of the pattern is maintained). To create each AM display, one of the eight stimulus types was embedded in an array of like inducers that were rotated so as not to form an enclosed region. (b). Examples of the fixed luminance (left panel) and highlight luminance (HL; right panel) conditions used in the experiments, shown for stimulus type 7. In the highlighted luminance condition, the AM inducers in each frame had a higher luminance than the others. This manipulation allowed observers to perform the task when the inducers' arrangement did not trigger global apparent motion.

to form a specific stimulus type. The remaining 6 inducers were rotated so that they did not produce an enclosed region. On successive frames, each inducer rotated ( $90^\circ$  or  $180^\circ$ ) in a manner that produced global AM of the four stimulus inducers (Figure 2b). We will refer to the set of four sequentially displaced inducers as the “AM inducers.” Individual inducers were spaced  $3^\circ$  apart (center-to-center). The speed of AM was controlled by varying the duration of each successive frame. Four different speeds were tested: 6.125, 11.25, 22.5, and 45 deg/s. [Movie 1](#) contains the AM stimuli for stimulus types 1, 7, and 8 ([click here to see demonstrations of the other stimulus types](#)).

From previous studies (Bravo et al., 1988; Seghier et al., 2000), it was known that observers would have difficulty perceiving AM for stimulus type 8 (no enclosed region). We therefore introduced a second condition in which the contrast of the AM inducers was increased relative to the other six inducers. We termed this the *highlight luminance* (HL) condition (Figure 3b). An HL condition was run for each of the eight stimulus types to evaluate the effect of highlighting in each case. The Weber contrast of the inducers compared with the background ( $(L_{\text{inducer}} - L_{\text{background}})/L_{\text{background}}$ ) was 2.54, that of the highlighted inducers was 4.27.

## Experimental design

In all, there were 8 stimulus types  $\times$  4 speeds  $\times$  2 conditions (fixed luminance and HL)  $\times$  2 directions of AM resulting in 128 different trial types. Trials were blocked

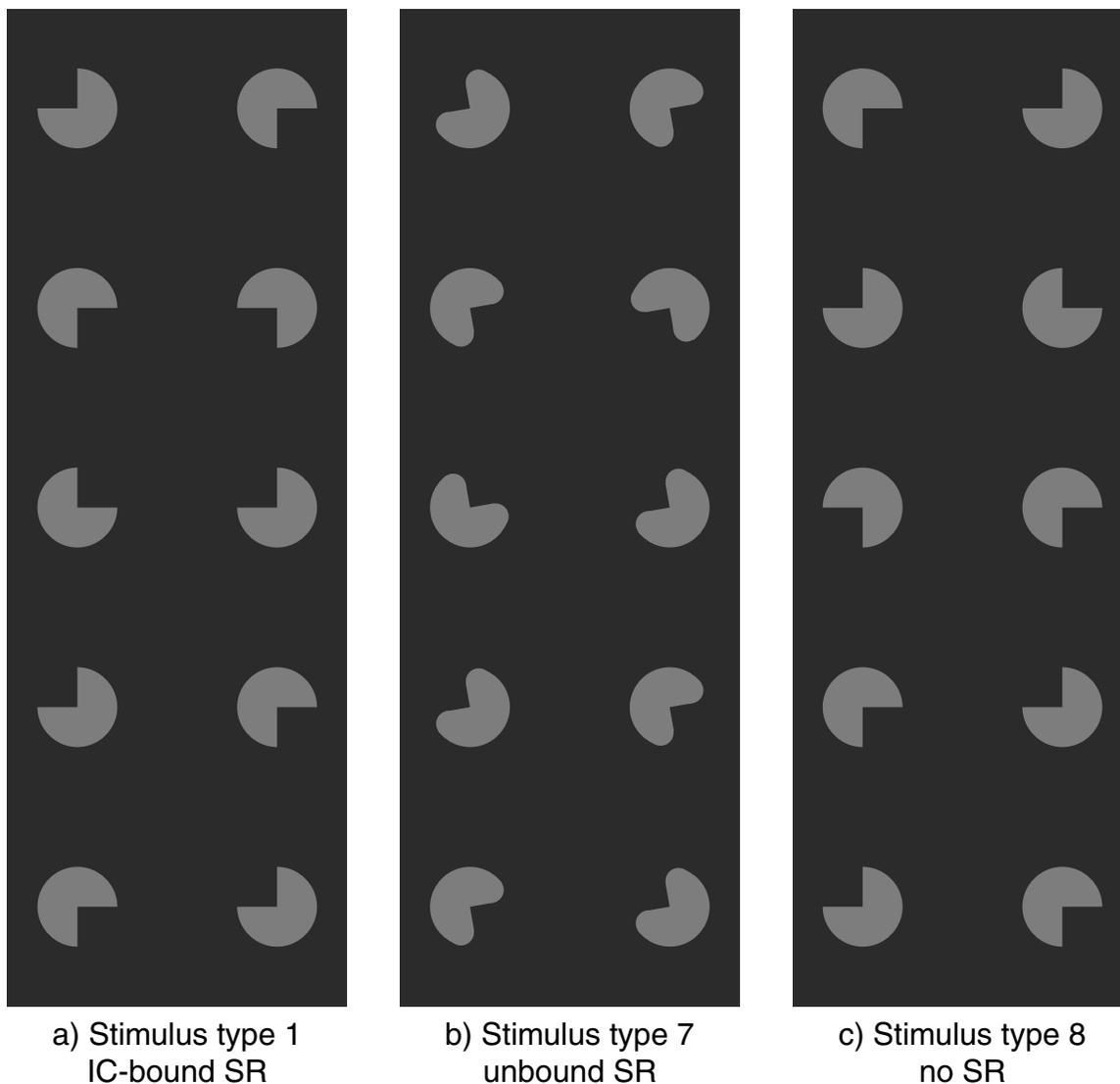
according to speed. Within a block, the 32 trial types were presented twice and randomly intermixed. The starting position of the AM inducers was counterbalanced (50% directly above fixation, 50% directly below) so observers could not use position as a cue for the direction of AM. Observers repeated each block twice. The order of block presentation was counterbalanced across observers using a  $4 \times 4$  Latin square. Observers fixated a cross ( $0.15^\circ$ ) presented at the center of the inducer array throughout each trial.

## Apparatus

Stimuli were displayed using Matlab and Psychtoolbox (Brainard, 1997; Pelli, 1997) on a Power Macintosh G4 computer with a 20-in. Sony Trinitron Multiscan 500PS screen (actual display width  $\times$  height =  $40 \times 30$  cm). The screen resolution was  $1600 \times 1200$  pixels, and the refresh rate was 75 Hz. Observers sat 57 cm from the screen, and a chin rest was used to reduce head motion.

## Data analysis

For each observer, mean response time (RT) and accuracy were calculated for each trial type (Condition  $\times$  Speed  $\times$  Stimulus Type). To enable comparison of RTs across different speeds, RTs were measured from the onset of the second frame (i.e., the onset of AM). The mean RT (correct trials only) across subjects was then



Movie 1. Movies demonstrating apparent motion for stimulus types 1 (a; IC-bound SR), 7 (b; unbound SR) and 8 (c; no SR). Apparent motion is seen when SRs are perceptually completed (stimulus types 1 and 7), but not when there is no SR present (stimulus type 8). The stimuli have the same spatial parameters used in the experiment when viewed from 57 cm away, on a 20-inch monitor with a screen resolution of 1024 x 768 pixels. The speed of apparent motion is 12 deg/sec (each frame of the sequence is displayed for 250 ms). Note that in the actual experiment the 4 AM inducers did not flash prior to the onset of apparent motion, this is only done here to help the reader identify their initial location and configuration. [Click on the link for demonstrations of the other stimulus types.](#)

calculated for each trial type and speed. Each observer's overall mean RT was subtracted from their data before calculating the standard error so that individual variations in mean RT do not effect error bars.

To examine closely the effects of stimulus type, speed, and condition as well as any interactions, we conducted a factorial analysis of variance (ANOVA) with these factors as predictors and  $\log(\text{RT})$  as the dependent variable. Examination of the residuals verified that the log conversion was adequate to satisfy the assumption of a normal distribution. Post hoc comparisons (Tukey HSD) were conducted to determine the contribution of individual stimulus types to the overall effects we found.

## Results and discussion

Overall, performance was markedly better for the stimuli that contained a salient region (SR; types 1–7) than for the stimulus with no SR (type 8). Figure 4 shows the results in terms of RTs (left axis) and error rates (right axis) for all eight stimulus types. The mean RT, averaged across all four speeds, for stimulus type 8 was 2.89 s. In contrast, for stimulus types 1–7, the mean RT collapsed across all speeds ranged from 0.60 to 1.29 s. The mean error rate for stimulus type 8 (32.1%) was higher than that for all other stimulus types (1.05–6.25%), indicating that there was no

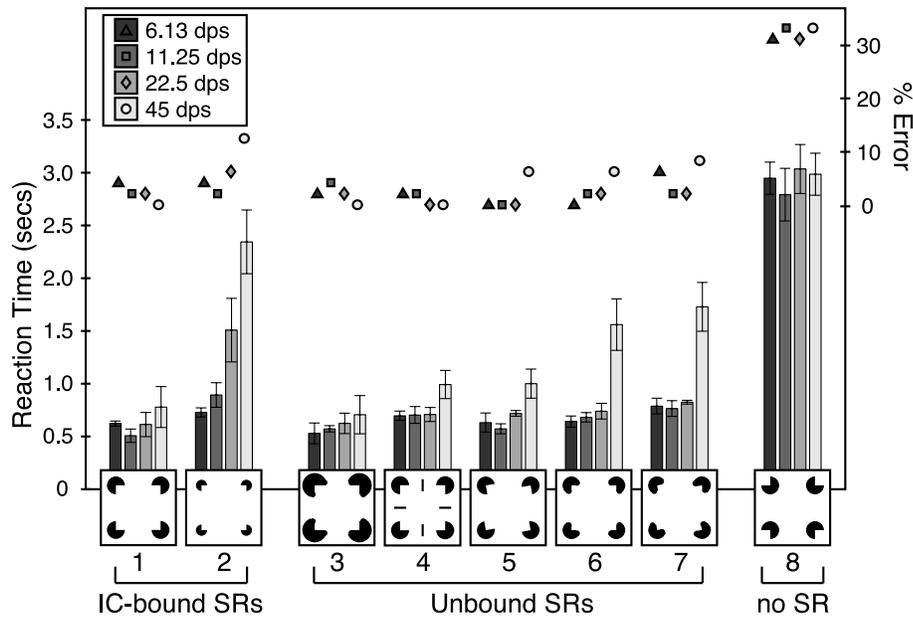


Figure 4. Mean RT and error rate for all eight stimulus types and speeds in the main experiment (fixed luminance). Above each stimulus type, mean RT (bars, scale shown on left; error bars indicate standard error across observers) and mean error rate (symbols, scale shown on right) are plotted for each of the four AM speeds. At all speeds, observers performed significantly better in terms of both RTs and error rates for stimulus types 1–7, which contained a salient region, than for stimulus type 8, which did not give rise to the perception of a salient region (far right).

speed–accuracy tradeoff. The good performance for stimulus types 1–7 indicates that the presence of an SR was sufficient to trigger AM in all cases. Thus, perceptually constructed regions do not need to correspond to an IC-bound surface in order for them to serve as matched items. Note that the better performance for stimulus types containing an SR (1–7) cannot be an effect of the global spatial configuration of the inducers, which was the same for all stimulus types (including type 8).

Highlighting the AM inducers with a different luminance level in the HL condition (Figure 3b) provided an additional correspondence cue for AM. To visualize the performance

advantage provided by the HL manipulation, we subtracted the across-observer mean RT for each stimulus type in the HL condition from the mean RT in the main experimental condition (fixed luminance). The results are shown in Figure 5. Performance for stimulus type 8 (no SR) was enhanced dramatically; RTs and error rates were reduced by an average (across speeds) of 1.3 s and 21.6%, respectively. This improvement was consistent with observer’s reports that in the HL condition of stimulus type 8 (no SR), they perceived apparent motion of the highlighted inducers, whereas for the fixed luminance stimuli they were basing their responses on the local rotational motion of the

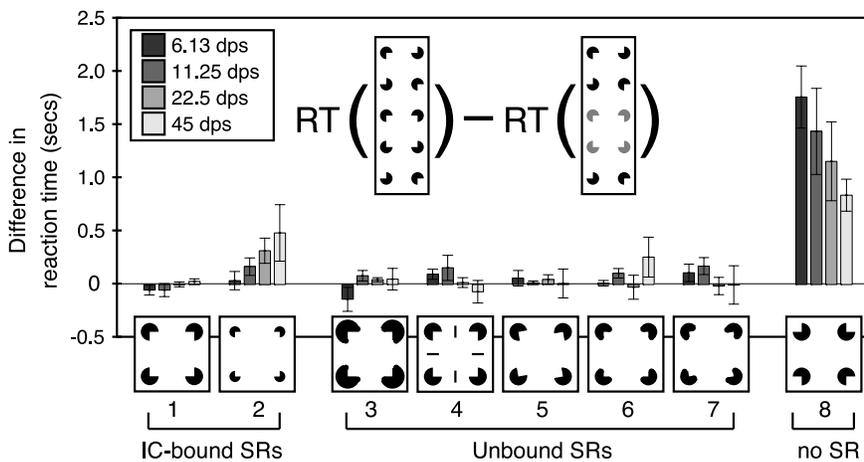


Figure 5. Mean RT benefit of highlighting the AM inducers (differences between the RTs for the fixed luminance stimuli and the HL stimuli). The largest benefit of the HL condition was seen for stimulus type 8 (far left). Comparatively, little to no benefit of highlighting was seen for stimulus types that contained a salient region (1–7).

inducers (which yielded above chance performance) because they did not perceive global AM. Despite the improvement provided by the added HL correspondence cue, performance for stimulus type 8 was still worse than for the stimuli that contained an SR (types 1–7). This poor performance, which indicates weak motion correspondence, has a number of potential sources. First, the highlighted inducers were much smaller than the perceived SRs, and since the distance traversed was the same, there was a larger relative gap between matched inducers. It is well known that increasing the displacement/size ratio of AM elements weakens their motion correspondence (essentially, it increases the perceived departure from continuous motion; Braddick, 1980; Kolers, 1972; Schechter & Hochstein, 1989; Shechter, Hochstein, & Hillman, 1988; Wertheimer, 1912). Second, there was more correspondence uncertainty for stimulus type 8, even in the HL condition. This is illustrated in Figure 6b; because of the presence of multiple HL inducers in each static frame, a highlighted inducer could be seen as “hopping” to the neighboring position, or to the next nearest neighbor, or remaining in place. In contrast, in stimulus types 1–7, each static frame contains only one SR, leaving no uncertainty in the matching (Figure 6a). Finally, even when motion is perceived for the AM inducers of stimulus type 8, it is only in the HL condition and thus is dependent upon the inducer color hopping from one inducer to the next (Figure 6b). Motion correspondence cues for a feature hopping from one surface to the next may not be as strong as when an entire region hops to a previously vacant location (Shechter & Hochstein, 1989).

Alongside the main finding—a marked difference in performance between stimulus types 1–7 and stimulus type 8—there were also some small differences in performance

within stimulus types 1–7. To investigate this more closely, we ran a full factorial ANOVA with  $\log(\text{RT})$  as the dependent variable (see Methods) and with stimulus type (1–7), condition (fixed luminance vs. HL), speed (four values), and observer (six, random factor) as predictors. Post hoc testing (Tukey HSD) was used to determine specific contributions to any significant main effect or interaction. The ANOVA revealed a main effect of stimulus type, ( $F(6, 2280) = 98.24, p < 10^{-17}$ ).

These differences in AM performance reveal differences in the degree of saliency among the SRs of stimulus types 1–7: All of them were perceptually salient—providing global correspondence features that led to good AM performance—but some (types 1, 3–5) were more salient than others (types 2, 6–7), leading to faster RTs. The most important thing to note about this effect is that differences in performance among stimulus types 1–7 did not follow the perceptual divide of presence versus absence of crisp bounding ICs. Performance was worst for stimulus type 2, a Kanisza square that has been shown to generate ICs despite its relatively low support ratio of 0.25 (Pillow & Rubin, 2002; Ringach & Shapley, 1996; Rubin, Nakayama, & Shapley, 1997; Shipley & Kellman, 1992). Post hoc testing indicated that performance on stimulus types 3 and 5 did not differ from performance on stimulus type 1, although previous research has shown that each of the manipulations for these stimulus types—eliminating the L junctions (type 3) and misaligning the inducing edges (type 5)—significantly reduces the perception of ICs compared with the intact stimulus type 1 (Kellman & Shipley, 1991; Rubin, 2001; Shipley & Kellman, 1990). This again suggests that whether an SR is bound by ICs plays little (if any) role in the detection of the enclosed region.

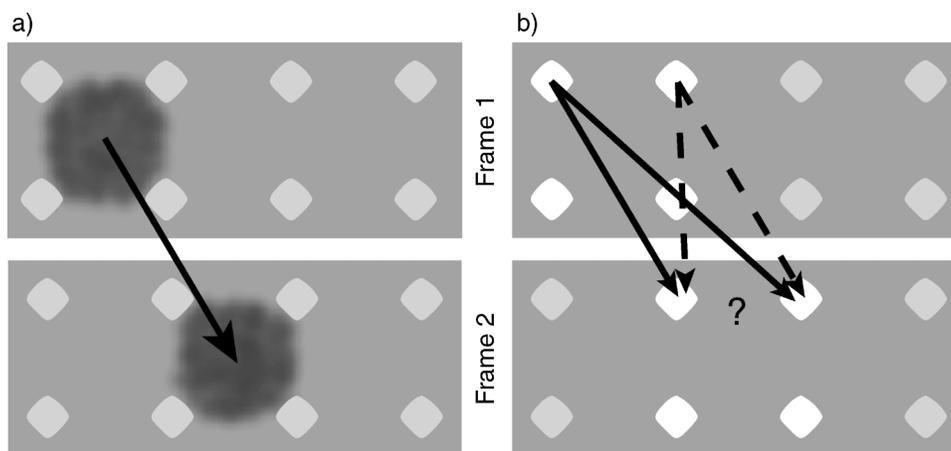


Figure 6. Illustration of perceived AM (indicated by the arrows between frames 1 and 2) for perceptually completed salient regions and for the highlighted AM inducers of stimulus type 8 (no global region). (a). AM of a globally completed region is unambiguous; observers perceive a region “hopping” from one position to the next. (b). In contrast, AM for the highlighted inducers of stimulus type 8 is ambiguous; observers could perceive inducers as “hopping” to the neighboring position, or the next nearest neighbor, or remaining in place. This ambiguity results in weaker motion correspondence. Note also that when there is no globally completed region (b), perceived AM relies on the inducer color “hopping” from one inducer to the next, whereas when there is a perceptually completed region (a), a whole region is seen hopping from one position to the next.

If the presence of bounding illusory contours does not determine performance in the AM task, is there another stimulus property that might? One possibility is that the amount of luminance-defined contour that is “missing” from the (putative) perimeter of the enclosed region plays a role, whether that region is perceptually bound by ICs or not. However, this conjecture would predict that performance on stimuli 2 and 6 would be similar—because in both cases luminance gradients are tangent to (“support”) 0.25 of the enclosed square’s perimeter—a prediction not borne out by the data. In fact, no completion process that relies solely on propagating signals along contours (Figure 1c) could predict the differences in performance between stimulus types 2 and 6. In contrast, region-based processes that mediate information not along contours but rather within regions (Figure 1d) may be sensitive to the marked overall difference in inducer size between the two stimulus types—which we will refer to as a difference in “region support.” In this study, we manipulate region support solely by varying the geometric shape of the inducers. However, other stimulus manipulations such as blurring or adding noise could also affect region support. At this stage, we do not offer an independent definition of region support. Instead, we propose an operational definition, based on performance in the AM paradigm used here (as well as possibly the visual search paradigm used by Gurnsey et al., 1996). As a starting point, our results offer a qualitative classification of “higher” (stimulus types 1, 3–7) versus “lower” (stimulus types 2, 8) region support. A more extensive exploration of the stimulus parameter space is needed to sufficiently constrain a general-purpose, quantitative definition. Such a definition could be used, in turn, to test the plausibility of different region-based computer vision algorithms (which can make different predictions on the effect of region support on segmentation) for human vision.

There was also a main effect of speed, ( $F(3, 2280) = 389.68, p < 10^{-17}$ ), as speed increased, performance worsened. Performance differences across speeds were not uniform across stimulus types, as revealed by an interaction of Speed  $\times$  Stimulus, ( $F(18, 2280) = 16.26, p < 10^{-17}$ ). At the lowest speed (6.13 deg/s), there were no significant differences in performance for stimulus types 1–7. At 11.25 deg/s, performance for stimulus 2 became significantly worse than that for stimulus types 1 and 3–5. As speed further increased, performance continued to worsen, more so for stimulus types 2, 6, and 7 than for types 1 and 3–5. Because in our study the speed of AM was determined by duration of each frame in the AM stimulus (i.e., the distance was fixed across all speeds), this speed dependency directly translates to a dependence on the “lifetime” of each AM frame: A 45 deg/s, each frame was presented for 67 ms, whereas at 6.13 deg/s, each frame was presented for 533 ms. This dependence on the duration of each frame suggests that the mechanisms responsible for SR detection proceed faster when there is more region

support. According to this interpretation, when stimuli with low region support are presented at high speeds (short presentation durations), the signal indicating the presence of an SR may not have time to reach full strength; thus, performance is selectively impaired. For stimulus types 4–7, performance at 45 deg/s is significantly worse than performance at the three lower speeds, indicating that the time it takes for the SR signal to reach full strength for these stimuli is approximately between 67 and 133 ms. Stimulus type 2 shows a significant impairment of performance at even longer presentation durations (22 deg/s). That AM was perceived for any of the SR stimuli at the highest speed indicates that the detection of SRs can be a very rapid process (<67 ms).

A main effect of condition, ( $F(1, 2280) = 30.17, p = 1.6 \times 10^{-7}$ ), revealed that performance was better in the HL condition than for the fixed luminance stimuli. This effect was not significant for individual stimuli with the exception of stimulus type 2, as revealed by a significant interaction of Condition  $\times$  Stimulus, ( $F(6, 2280) = 2.96, p = 0.007$ ). This benefit of the HL condition is another indicator that stimulus type 2 contained only weak region support for an SR. In contrast, the HL condition did not provide a benefit for stimulus types 1 and 3–7 because the SR was strong enough to serve as robust cue for motion correspondence.

Finally, there was also a 3-way interaction of Observer  $\times$  Speed  $\times$  Stimulus, ( $F(90, 2280) = 3.091, p < 10^{-17}$ ); for all observers, performance was significantly impaired for stimulus type 2 compared with stimulus 1, but the degree of impairment for stimulus types 4–7 varied across observer, with some showing dramatic impairments and others showing only trends of worsening performance. This three-way interaction was also reflected in significant two-way interactions of Observer  $\times$  Speed, ( $F(15, 2280) = 37.50, p < 10^{-17}$ ), and Observer  $\times$  Stimulus, ( $F(30, 2280) = 6.08, p < 10^{-17}$ ), and in the main effect of observer, ( $F(5, 2280) = 546.34, p < 10^{-17}$ ).

## General discussion

We used an apparent motion (AM) paradigm to probe the processes underlying perceptual completion. The displays were modified versions of a stimulus introduced by Ramachandran (1985, 1986), where the orientation of “pacman”-shaped local inducers was varied between successive frames to create translational AM of an illusory Kanizsa square (Figure 2a and 2b; see also Bravo et al., 1988; Goebel, Khorrám-Sefat, Muckli, Hacker, & Singer, 1998; Seghier et al., 2000). In our stimuli, the shape and/or alignment of the inducers were modified in ways that eliminated the bounding illusory contours (Kellman & Shipley, 1991; Rubín, 2001; Shipley & Kellman, 1990). Although ICs are no longer perceived, the stimuli still

give rise to a clear impression of an enclosed region (Figure 3a, stimulus types 3–7). Borrowing from the computer vision literature, we refer to these perceptually completed regions as ‘salient.’ Our results indicate that SRs can give rise to robust AM even when they are not bounded by ICs: Under a wide range of parameters, performance in a motion direction discrimination task was as good or better for stimulus types 3–7 (SRs not bounded by ICs) than for stimulus types 1–2 (IC-bound SRs).

The present study joins two other studies that have called for the reinterpretation of previous experiments involving Kanizsa-type illusory figure stimuli. Gurnsey et al. (1996) showed that the rapid search performance of Kanizsa figures, originally reported by Davis and Driver (1994), was maintained when the stimuli were modified to eliminate the bounding ICs (e.g., one of their search targets resembled our stimulus type 4, Figure 3a). In an fMRI study, Stanley and Rubin (2003) found that LOC responses to Kanizsa-type illusory figures, previously reported by Hirsch et al. (1995) and Mendola et al. (1999), were again maintained under stimulus manipulations that eliminated the bounding ICs but retained the enclosed SR. How are we to interpret these results? Before we turn to answer this, we believe it is important to comment on one reading of our results that we do not endorse. The main message of the present study, as well as the two others mentioned above (Gurnsey et al., 1996; Stanley & Rubin, 2003), is not one of dismissing the original studies that they followed up. The original observations about illusory figures—that they can undergo apparent motion, “pop out” in search displays and activate the LOC—remain important despite the later finding that bounding ICs are not necessary for these phenomena to occur. This is because all along, the importance of these observations lay not in what they taught us about ICs per se but in what we can learn from them about scene segmentation and perceptual completion.

We interpret our results within a theoretical framework that makes a distinction between contour-based and region-based segmentation processes and posits that they play complementary roles in human visual scene segmentation. The concept of region-based segmentation arose in the field of computer vision, in the context of algorithms for rapid (if somewhat crude) segmentation of cluttered real-world images. In the past, computer vision scientists put emphasis on segmenting images based on the output of edge-detection filters (e.g., Marr, 1982). This approach, which we refer to as contour-based, was consistent with (and quite likely inspired by) physiological findings that early visual cortical cells respond selectively to luminance edges. However, it proved limited, in large part because of the need to perform contour-completion computations (to overcome occlusion, shadows, noise, etc.), which are resource-intensive and slow. This led computer vision scientists to try a different approach, that is, going *from the surface to its boundaries*, rather than the other way

around. The goal of region-based algorithms is to identify the regions of highest saliency in the image: contiguous sets of pixels that likely correspond to major objects in the scene. For this purpose, edges are not the only source of useful information: For example, knowledge that an image region is uniform in luminance/color/texture—or, more likely, that the variation in those properties within the region is significantly smaller than the variation between regions—is just as important. Consequently, region-based algorithms benefit from the possibility of propagating signals in all directions (rather than only along contours), which can speed up convergence (cf. Pao, Geiger, & Rubin, 1999; Sharon et al., 2000; Shi & Malik, 2000). Another advantage is that region-based computations can be performed at multiple image scales and sped-up by interaction between finer scales and progressively “coarsened” resolutions (Sharon et al., 2000). At the same time, region-based processes alone provide only a crude segmentation and can yield “false alarms”—when a region, which, in fact, belongs to the background, is classified as salient (for examples from the output of two leading algorithms on real-world images, see Figure S4 in Supplemental Data of Stanley & Rubin, 2003, <http://www.neuron.org/cgi/content/full/37/2/323/DC1>).

The strategy that has emerged in computer vision as particularly useful for performing segmentation of real-world images in realistic time scales is therefore to combine the two types of processes. Computationally intensive contour-based processing is restricted to select image regions, those identified as “salient” by rapid region-based parsing. Depending on whether the contour and junction analysis support the status of a region as a figural surface or not, further iterative cross-talk between the two types of processes may take place. (In addition, some or all of the high-saliency regions can be fed directly into upstream modules that perform image-based “template-matching” object recognition; cf. Ullman, 1996.)

Stanley and Rubin (2003) suggested that the brain might be employing a similar strategy to speed up scene segmentation. Based on their finding that the human lateral occipital complex (LOC) shows elevated fMRI activity in response to SRs even when they are not bound by ICs (i.e., when they do not correspond to an actual surface in the underlying scene), they suggested that the LOC is involved in the type of crude-but-fast region-based processing that gives rise to the detection of SRs. They further suggested that the LOC might be well suited for performing such operations because of its large receptive fields. Finally, they proposed that the LOC might be directing, via feedback processes, earlier visual areas (V1/V2) to perform more detailed contour-completion computations on candidate surfaces. This idea is supported by electroencephalogram and magnetoencephalogram data showing that there is a very early (88- to 100-ms onset; Halgren, Mendola, Chong, & Dale, 2003; Murray et al., 2002) LOC response to Kanizsa-type figures, followed by

activity in early visual cortex. As in the case of the computer vision algorithms, this hypothesized process most likely does not occur in one step and would benefit from iterative reinforcement between early visual areas and the LOC.

The two-way processing stream described here bears some resemblance to that proposed by Hochstein and Ahissar (2002) in their reverse hierarchy theory. There, too, it was proposed that a fast-but-crude representation of the scene is performed by higher visual cortical areas, whereas a more detailed analysis of the image (e.g., to find an odd-shaped target in a search display) requires the involvement of early cortical regions. It is encouraging that theories developed to explain quite different aspects of human vision (segmentation in our case, perceptual learning of visual search in Hochstein and Ahissar's case) and that rely on entirely non-overlapping bodies of data reach such converging conclusions. At the same time, there is one notable difference between the two theories. In reverse hierarchy theory, the detailed processing in early cortex occurs as a consequence of conscious effort of the observer (as manifested by the label “vision with scrutiny”). In contrast, the feedback to the early cortex, which, as we hypothesize, directs contour-based processing to select image regions, is stimulus-driven and automatic. Indeed, the outcomes of these processes—such as the perception of ICs—occur without need for observers' conscious effort.

The current study supplements our previous work (Stanley & Rubin, 2003) by providing behavioral evidence for the perceptual validity of SRs. This, in turn, implies the operation of region-based processes that give rise to SRs in the image. Furthermore, the results suggest that apparent motion of SRs can be used as a tool to quantify their strength and how rapidly they are detected. This method and other indicators—such as whether an image region “pops out” in a search display (Gurnsey et al., 1996)—can be used to further investigate the underlying mechanisms of SR detection. An important direction for future research is to determine what image cues launch region-based completion mechanisms and how the detection and integration of these cues are implemented in the brain. This would benefit greatly from a dialog between computer vision researchers, who have identified many image properties (e.g., brightness, texture, size) useful for performing an initial crude segmentation of a scene, and experimentalists, who can determine which image cues play a role in the perception and neural representation of salient regions.

## Acknowledgments

Funding was provided by the National Institutes of Health's Ruth L. Kirschstein National Research Service Award predoctoral fellowship F31 MH65805-01A2, the National Eye Institute grant EY14030, the Sloan Foundation, and support from the Beatrice and Samuel A. Seaver

Foundation to New York University. We thank Davi Geiger and Eitan Sharon for helpful discussions.

Commercial relationships: none.

Corresponding author: Damian A. Stanley.

Email: das@cns.nyu.edu

Address: The Center for Neural Science, New York University, 4 Washington Place, Room 809, New York, NY 10003.

## References

- Braddick, O. J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 290(1038), 137–151. [PubMed]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. [PubMed]
- Bravo, M., Blake, R., & Morrison, S. (1988). Cats see subjective contours. *Vision Research*, 28(8), 861–865. [PubMed]
- Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. [see comment]. *Nature*, 371(6500), 791–793. [PubMed]
- Goebel, R., Khorrám-Sefat, D., Muckli, L., Hacker, H., & Singer, W. (1998). The constructive nature of vision: Direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal of Neuroscience*, 10(5), 1563–1573. [PubMed]
- Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, 13(2), 159–166. [PubMed]
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10–11), 1409–1422. [PubMed]
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92(2), 173–211. [PubMed]
- Gurnsey, R., Poirier, F. J., & Gascon, E. (1996). There is no evidence that Kanizsa-type subjective contours can be detected in parallel. *Perception*, 25(7), 861–874. [PubMed]
- Guzman, A. (1969). Decomposition of a visual scene into three-dimensional bodies. In A. Grasselli (Ed.), *Automatic interpretation and classification of images* (pp. 243–276). New York: Academic Press.

- 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. [[PubMed](#)]
- Heitger, F., Rosenthaler, L., von der Heydt, R., Peterhans, E., & Kubler, O. (1992). Simulation of neural contour mechanisms: From simple to end-stopped cells. *Vision Research*, 32(5), 963–981. [[PubMed](#)]
- Hirsch, J., DeLaPaz, R. L., Relkin, N. R., Victor, J., Kim, K., Li, T., et al. (1995). Illusory contours activate specific regions in human visual cortex: Evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 92(14), 6469–6473. [[PubMed](#)] [[Article](#)]
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791–804. [[PubMed](#)]
- Kanizsa, G. (1987). Margini quasi-percettivi in campi con stimolazione omogenea. English translation in S. Petry & G. Meyer (Eds.), *The perception of illusory contours* (pp. 40–49). New York: Springer-Verlag. (Original work published in 1955)
- Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23(2), 141–221. [[PubMed](#)]
- Kolers, P. A. (1972). *Aspects of motion perception*. New York: Pergamon Press.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92(18), 8135–8139. [[PubMed](#)] [[Article](#)]
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.
- Mendola, J. D., Dale, A. M., Fischl, B., Liu, A. K., & Tootell, R. B. (1999). The representation of illusory and real contours in human cortical visual areas revealed by functional magnetic resonance imaging. *Journal of Neuroscience*, 19(19), 8560–8572. [[PubMed](#)] [[Article](#)]
- Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 125–152). Cambridge, MA: MIT Press.
- Murray, M. M., Wylie, G. R., Higgins, B. A., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). The spatiotemporal dynamics of illusory contour processing: Combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. *Journal of Neuroscience*, 22(12), 5055–5073. [[PubMed](#)] [[Article](#)]
- Nieder, A. (2002). Seeing more than meets the eye: Processing of illusory contours in animals. *Journal of Comparative Physiology. A, Sensory, Neural, and Behavioral Physiology*, 188(4), 249–260. [[PubMed](#)]
- Pao, H., Geiger, D., & Rubin, N. (1999). *Measuring convexity for figure/ground separation*. Paper presented at the 7th IEEE International Conference on Computer Vision, Corfu, Greece.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. [[PubMed](#)]
- Pillow, J., & Rubin, N. (2002). Perceptual completion across the vertical meridian and the role of early visual cortex [see comment]. *Neuron*, 33(5), 805–813. [[PubMed](#)]
- Ramachandran, V. S. (1985). Apparent motion of subjective surfaces. *Perception*, 14(2), 127–134. [[PubMed](#)]
- Ramachandran, V. S. (1986). Capture of stereopsis and apparent motion by illusory contours. *Perception & Psychophysics*, 39(5), 361–373. [[PubMed](#)]
- Ringach, D. L., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research*, 36(19), 3037–3050. [[PubMed](#)]
- Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception*, 30(3), 339–366. [[PubMed](#)]
- Rubin, N., Nakayama, K., & Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Current Biology*, 7(7), 461–467. [[PubMed](#)]
- Seghier, M., Dojat, M., Delon-Martin, C., Rubin, C., Warnking, J., Segebarth, C., et al. (2000). Moving illusory contours activate primary visual cortex: An fMRI study. *Cerebral Cortex*, 10(7), 663–670. [[PubMed](#)] [[Article](#)]
- Sharon, E., Brandt, A., & Basri, R. (2000). *Fast multiscale image segmentation*. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC.
- Shechter, S., & Hochstein, S. (1989). Size, flux and luminance effects in the apparent motion correspondence process. *Vision Research*, 29(5), 579–591. [[PubMed](#)]
- Shechter, S., Hochstein, S., & Hillman, P. (1988). Shape similarity and distance disparity as apparent motion correspondence cues. *Vision Research*, 28(9), 1013–1021. [[PubMed](#)]
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888–905.

- Shipley, T. F., & Kellman, P. J. (1990). The role of discontinuities in the perception of subjective figures. *Perception & Psychophysics*, 48(3), 259–270. [[PubMed](#)]
- Shipley, T. F., & Kellman, P. J. (1992). Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception & Psychophysics*, 52(1), 97–106. [[PubMed](#)]
- Spillmann, L., & Dresch, B. (1995). Phenomena of illusory form: Can we bridge the gap between levels of explanation? *Perception*, 24(11), 1333–1364. [[PubMed](#)]
- Stanley, D. A., & Rubin, N. (2003). fMRI activation in response to illusory contours and salient regions in the human lateral occipital complex. *Neuron*, 37(2), 323–331. [[PubMed](#)]
- Ullman, S. (1996). *High-level vision: Object recognition and visual cognition* (Hardcover ed.). Cambridge, MA: MIT Press.
- Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 61, 161–265.