

Contour discontinuities subserve two types of form analysis that underlie motion processing

Peter Ulric Tse* and Gideon P. Caplovitz

H.B. 6207, Moore Hall, Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA

Abstract: Form analysis subserves motion processing in at least two ways: first, in terms of figural segmentation dedicated to solving the problem of figure-to-figure matching over time, and second, in terms of defining trackable features whose unambiguous motion signals can be generalized to ambiguously moving portions of an object. The former is a primarily ventral process involving the lateral occipital complex and also retinotopic areas such as V2 and V4, and the latter is a dorsal process involving V3A. Contour discontinuities, such as corners, deep concavities, maxima of positive curvature, junctions, and terminators, play a central role in both types of form analysis. Transformational apparent motion will be discussed in the context of figural segmentation and matching, and rotational motion in the context of trackable features. In both cases the analysis of form must proceed in parallel with the analysis of motion, in order to constrain the ongoing analysis of motion.

Keywords: Motion; Contour discontinuities; Form; Vision.

The importance of contour discontinuities

Contour discontinuities are more informative about object shape and motion than other portions of contour. They function in part as invariants in the image that permit shape and motion to be extracted and constructed from information available in the visual array. One of the first to realize the importance of contour discontinuities was Attneave (1954). He realized that the visual system is more sensitive to the presence of local maxima of positive or negative contour curvature (CC) than it is to inflection points. Attneave argued on the basis of information theory that contour discontinuities such as highly curved portions of a contour provide more information than straight sections of contour because the location of neighboring points is more unpredictable

around contour discontinuities than straight portions of contour. This insight was extended by Biederman (1987), who showed that objects are unrecognizable if contour discontinuities are removed, but are easily recognizable if an equal amount of contour is removed that spares contour discontinuities. Neuropsychological work has also demonstrated the importance of contour discontinuities. Humphreys et al. (1994) found that Balint's patients, who can only perceive one object at a time, tended to perceive the version of a square or diamond defined by corner segments of contour over those defined by the straight sections of contour between corners.

Because contour discontinuities are so informative, it is not surprising that classes of detectors have evolved to detect them. Fast search rates in the visual search paradigm have generally been regarded as evidence that the visual system is predisposed to rapidly detect a given image feature, event, or configuration. It is well known that

*Corresponding author. Tel.: +1-603-646-4014;
E-mail: Peter.U.Tse@dartmouth.edu

tangent and curvature “pop out” (i.e., are rapidly detected, as determined by flat search slopes as a function of distractor set size) when the distractors are straight lines (Treisman and Gormican, 1988; Wolfe et al., 1992; see also Zucker et al., 1989). Kristjansson and Tse (2001) showed that CC discontinuities pop out among continuous curvature distractors. In contrast, straight lines do not pop out among curves, and continuous curves do not pop out among discontinuous ones. The search asymmetry for curved targets among straight-line distractors attests to the visual system’s bias to preferentially detect information about curvature rapidly. Similar results using other contour discontinuities have led to the conclusion that contour discontinuities in general, including tangent discontinuities, junctions, corners, deep concavities, terminators, curvature, and even CC discontinuities, are basic features to which the visual system is highly sensitive.

Contour discontinuities and figural matching

Here we will focus on transformational apparent motion (TAM), because it is a phenomenon that makes apparent the importance of form processing in determining perceived motion. When an object discretely and instantaneously changes its shape, observers typically do not perceive the abrupt transition between shapes that in fact occurs. Rather, a continuous shape change is perceived. Although TAM is a faulty construction of the visual system, it is not arbitrary. From the many possible shape changes that could have been inferred, usually just one is perceived because only one is consistent with the shape-based rules that the visual system uses to (1) segment figures from one another within a scene, and (2) match figures to themselves across successive scenes. TAM requires an interaction between neuronal circuits that process form relationships with circuits that compute motion trajectories. In particular, this form–motion interaction must happen before TAM is perceived, because the direction of perceived motion is dictated by form relationships among figures in successive images. Overall, the occurrence of TAM is consistent with a view of motion processing

comprised of two subsystems: one low level, driven by the processing of motion-energy extracted from changes in various features, and the other high level, driven by global contour-based form analysis and the attentional tracking of figures.

Historical background to TAM

In recent decades, motion perception has predominantly been studied from the perspective of neurophysiology, occasioned by the dramatic discovery of motion-sensitive neurons (e.g., Barlow and Levick, 1965). Psychophysical and physiological observations support the notion that motion perception arises from the responses of populations of neurons tuned to various spatiotemporal offsets. Recent formal models of motion perception emphasize “motion-energy,” which could be captured by the receptive fields of such neurons (e.g., Hassenstein and Reichardt, 1956; Reichardt, 1957, 1961; Barlow and Levick, 1965; Adelson and Bergen, 1985; van Santen and Sperling, 1985). According to such motion-energy detection models, a given motion-sensitive neural “comparator” signals motion when it registers some feature, most commonly relative luminance, at one location, and then later registers that feature at a second location. A simplified low-level motion unit has two receptive fields with a particular spatial offset, and their activity is compared upon a given temporal offset (e.g., a “Reichardt detector”; Hassenstein and Reichardt, 1956). If the two receptive regions become successively active in the preferred order, the unit responds, signaling that something has moved in a direction and speed consistent with the spatiotemporal offset that the unit is wired up to detect. Such “low-level” motion processing does not segment the image, but rather decomposes the image pattern into components of motion-energy, allowing separate responses of many different comparators to any one location at the same time. This property of registering the motion-energy of multiple, superimposed components at a location, rather than treating the pattern at that location as a single entity, is the critical difference between low-level and high-level motion-processing systems.

High-level motion processing, in contrast, has been described as a process of identifying forms and then matching those forms across time intervals (Anstis, 1980; Braddick, 1980). Grouping and segmentation procedures would treat a pattern at a single location as a single entity that is not decomposable into Fourier components. The high-level motion-processing system would not detect spatiotemporal displacements of luminance or any other image feature per se. Rather, the high-level system would track a single thing as it moved, even when the features defining that thing changed, and, indeed, would do so even if motion-energy implied motion in a direction opposite that of the tracked figure.

Motivated by early speculation about the possibility of a form-sensitive motion-processing stream, researchers went in search of form-motion interactions but were unable to demonstrate strong effects of form on motion. What studies using translational apparent motion have generally shown is that few of the elements of scene or form analysis have much impact on matching in apparent motion (see e.g., Kolers and Pomerantz, 1971; Kolers and von Grünau, 1976; Navon, 1976; Burt and Sperling, 1981; Baro and Levinson, 1988; Cavanagh et al., 1989; Victor and Conte, 1990; Dawson, 1991). Indeed, low spatial frequencies appear to count for more than shape (Ramachandran et al., 1983), and the most potent factor is found to be spatiotemporal proximity. There are some examples of object properties — shape, color, depth — playing a minor role in matches (e.g., Green, 1986a, b, 1989). But the effect of these factors is only revealed when the much stronger factor of proximity has been carefully controlled for. This tendency to match elements on the basis of proximity in successive displays has been called “the nearest neighbor principle” (Ullman, 1979). Relative to proximity factors, color and form factors seem insignificant in determining matches between scenes using translational apparent motion as a probe. The conclusion that form played little role in matching elements in apparent motion displays seemed to undermine the defining characteristic of high-level motion processing, namely that form extraction precedes motion perception. Results were not that

different from what would have been predicted on the basis of low-level mechanisms.

If versions of the standard low-level motion units were available with very large receptive fields, they could respond to isolated items of two frames even over large spatial separations. However, they could not resolve the details of each item’s shape. They would signal motion of the nearest neighbor pairs and these simple receptive fields would ignore the form and color details of the items. There are examples of apparent motion that deviate from the expected properties of low-level detectors, such as matching opposite-contrast items or matching across eyes, but these examples do not rule out contributions from low-level units in most apparent motion displays. They only suggest that low-level mechanisms cannot be solely responsible for all motion phenomena.

The seeming indifference of motion processing to form information in apparent motion meshed well with neurophysiological findings that the low-spatial, high-temporal resolution magnocellular system, specialized for the processing of motion and spatial relationships, responds poorly to contours and boundaries defined only by color contrast (e.g., Ramachandran and Gregory, 1978; Livingstone and Hubel, 1987), whereas the high-spatial, low-temporal resolution parvocellular/interblob system, specialized for form analysis responds poorly to motion (e.g., Livingstone and Hubel, 1987; Schiller, 1991).

Because both psychophysical and neurophysiological evidence seemed to indicate that motion analysis was not dependent on form-based matching, many in the visual neuroscience community reached the conclusion by the early 1990s that the presumed stage of form extraction was either not present or not significant in high-level motion processing. Many researchers in the recent past, at least by implication, therefore attempted to reduce the high-level motion-processing stream to a variant of the low-level motion-processing stream. For example, instances of motion phenomena that were not easily reducible to luminance-defined motion-energy processing were deemed to be processed by detectors that had additional properties, such as rectification prior to the stage where motion-energy would be detected by more or less

traditional motion-energy detectors. A common claim was that most types of image motion could be detected and analyzed using elaborations of motion-energy detectors that were not explicitly tuned to form or contour relationships. (See, e.g., Braddick, 1974; Marr and Ullman, 1981; Adelson and Movshon, 1982; Watson and Ahumada, 1983; van Santen and Sperling, 1984; Adelson and Bergen, 1985; Burr et al., 1986; Chubb and Sperling, 1988; Cavanagh and Mather, 1989; Lu and Sperling, 1995a, 2001.) The emergence of TAM in the past 10 years has, however, changed this debate by offering a reassessment of the role that form analysis plays in high-level motion processing, and by calling into question recent attempts to reduce all motion processing to the analysis of various orders of motion-energy processing.

Transformational apparent motion

Over the past several years authors have investigated a new type of motion phenomenon that occurs when two spatially overlapping shapes presented discretely in time appear to transform smoothly and illusorily from the first shape into the second as if the sequence were animated using a succession of intermediate shapes, as shown in Fig. 1. The illusion of apparently smooth and continuous shape change was termed TAM (Tse et al., 1998) in order to contrast it with the translational apparent motion first discovered by Gestalt psychologists (e.g., Wertheimer, 1912; Kenkel, 1913). TAM occurs when an object is abruptly flashed on next to an abutting static object, causing the new object to appear to smoothly extend from the static object. Upon offset, it appears to smoothly retract into the static object. Abutment or near abutment appears to be a necessary condition for the occurrence of TAM. A precedent to TAM was first described by Kanizsa (1951, 1971) and termed “polarized gamma motion.” This phenomenon was rediscovered in a more compelling form by Hikosaka et al. (1991, 1993a, b). They showed that when a horizontal bar is presented shortly after a cue, the bar appears to shoot away from the cue. Calling this phenomenon “illusory line motion,” they hypothesized that the

effect was due to the formation of an attentional gradient around the cue. Because an attentional gradient would fall off with distance from the cue, and because it has been shown that attention increases the speed of stimulus detection (Titchener, 1908; Sternberg and Knoll, 1973; Stelmach and Herdman, 1991, 1994), Hikosaka et al. hypothesized that illusory line motion occurs because of the asynchronous arrival of visual input to a motion detector such as human area hMT+. Shortly after, other authors (e.g., von Grünau and Faubert, 1994; Faubert and von Grünau, 1995) discovered low-level contributions to this phenomenon that could not easily be attributed to a gradient of attention. For example, a red line flashed on all at once between two existing spots, one red and one green, would always appear to move away from the same color spot, regardless of how attention was allocated. Soon after this, other authors (Downing and Treisman, 1995, 1997; Tse and Cavanagh, 1995; Tse et al., 1996, 1998) rejected the attentional gradient account and suggested that this phenomenon was actually a type of apparent motion where the initial shape appears to smoothly transform into the second shape. These authors have shown that TAM arises even when attention is paid to the opposite end of the cued location, implying that there must be other contributors to the motion percept than a gradient of attention. The attentional gradient account has now been disproved in direct psychophysical tests (Hsieh et al., 2005), meaning that an alternative explanation of TAM is required.

Tse and colleagues (Tse et al., 1998; Tse and Logothetis, 2002) have shown that figural parsing plays an essential role in determining the perceived direction of TAM. Figural parsing involves a comparison of contour and surface relationships among successive scenes. In particular, the visual system appears to infer which figures at time 2 are derived from which figures at time 1 on the basis of contour and surface relationships. If a given figure has a different shape at times 1 and 2, a continuous deformation between those shapes is constructed and perceived, presumably because the new figure is inferred to be a change in the shape of an already existing figure. The key point here is that form processing must temporally precede or at the

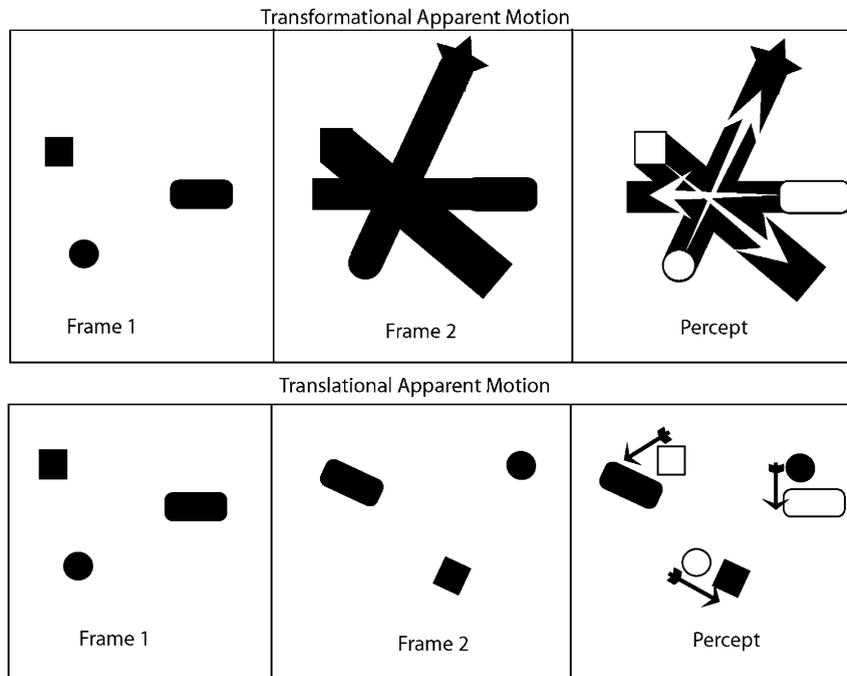


Fig. 1. Transformational apparent motion (TAM) (top row) is contrasted with translational apparent motion (bottom row). When the figures in frame 1 (leftmost column) are replaced instantaneously with the figures in frame 2 (middle column), then a percept (rightmost column) of TAM (i.e., smooth deformations) results when figures spatiotemporally abut, but of rigid translations results when they do not abut. Note that even though frame 2 may consist of a single contiguous shape, as in the top middle frame, multiple motions can be perceived to occur simultaneously, as depicted by the arrows on the top right.

very least accompany the motion processing that subserves the percept of TAM. This is because the motion that is seen depends on how figures at time 1 have been matched to figures at time 2. Determining that something has moved requires that the “something” be identified in the first instant and then paired off with what is presumed to be the same thing in the next instant. The first component of this processing is to identify candidates at both instants, and the second is to match them. We can call the first component a parsing step and the second component a matching step. By “parsing” is meant the spatial isolation and specification of individual objects, including any necessary segmentation away from overlapping objects or background elements of the image, as well as the discounting of noise.

A representative example of TAM is shown in Fig. 1. TAM can occur behind partially occluding objects and over illusory surfaces that have not been cued to attract attention to one location over

another, and TAM can appear to emerge out of the picture plane, implying that it can take place over 3D representations after a stage of modal and amodal completion (Tse and Logothetis, 2002). Whereas translational apparent motion is generally insensitive to shape and color constraints so long as the two stimuli presented remain within the optimal range of spatiotemporal offsets (see, e.g., Cavanagh et al., 1976), TAM is sensitive to such shape and color constraints, because these can be used by the parser to disambiguate figures in scenes that can only be ambiguously parsed otherwise. Moreover, TAM can occur over figures that have been defined by changes in successive subregions of a field of random dots (Hsieh and Tse, 2006).

TAM reveals that parsed figures are matched to parsed figures from scene to scene in high-level motion processing, in some cases violating nearest-neighbor principles. In TAM, new image data appearing closer to one figure than another, can still

get matched as comprising a shape change in the more distant figure. Tse and colleagues (Tse et al., 1998; Tse and Logothetis, 2002) showed that a set of parsing and matching principles based on analyzing contour relationships among abutting and successive figures aids in determining figural identity within and between scenes holds for TAM. These principles approximately reduce to the nearest-neighbor principle for cases of translational apparent motion. Their behavioral research to date demonstrates that a stage of figural parsing and matching precedes the perception of motion in TAM displays. This stage of segmentation takes place on the basis of Gestalt-like grouping principles that process the spatiotemporal relationships among figures. Figure formation and matching appear to primarily take place on the basis of good contour continuity (indicating figural identity between scenes) and contour discontinuity (indicating possible figural difference). In particular, the high-level motion-processing system receives segmented and completed contours, and perhaps surfaces as input. The percept of motion is therefore in part the perception of how figures have translated and transformed their shape.

The importance of figural parsing could not have been revealed by research into translational apparent motion, because in translational apparent motion displays, the parsing of each successive scene is generally given unambiguously in the sense that figures are spatially distinct. In translational apparent motion displays, a figure seems to disappear at one location and reappear at a different, nonoverlapping location some time later. The problem in translational apparent motion experiments has generally been the match between figures, not the parsing of figures. However, in TAM displays, there is usually ambiguity in determining which figure in one scene has become which figure in the following scene, because of the spatiotemporal overlap of succeeding figures. That is, in the case of the apparent shape transformations of figures, new image data generally appear without the disappearance of the figure (s) that existed in the previous scene (although brief “figureless” intervals are tolerated, as in translational apparent motion). In TAM, the parsing problem has to be solved before the problem of matching or

maintaining figural identity across successive scenes can be attempted. Since the image itself is not parsed, the visual system faces a problem of ambiguity in its efforts to correctly parse the image so as to coincide with the actual segmentation of the world into independent, but abutting or overlapping, objects. Since many possible parsings are consistent with a single image, the visual system has evolved default processes for solving the parsing problem, realized, in part by contour-based completion and segmentation of figures. This contour-based completion and segmentation is driven in large part by an analysis of contour continuity and contour discontinuity relationships among successive, abutting, or overlapping figures.

The neuronal correlates of TAM

Recent functional magnetic resonance imaging (fMRI) work examined the neuronal correlates of TAM (Tse, 2006). The blood oxygen level dependent (BOLD) signal in several cortical areas was found to increase during the perception of TAM relative to the perception of control stimuli in which TAM was not observed. In particular, a region of interest analysis revealed that the BOLD signal in areas V1, V2, V3, V4, V3A/B, hMT+ (the human homologue of macaque MT) and LOC (lateral occipital complex) increased during the perception of TAM relative to the control. An additional, whole brain analysis revealed an additional area in the posterior fusiform gyrus that was also more active during the percept of TAM than during control. We can therefore conclude that the neural basis of TAM resides in these and perhaps other areas. Because TAM is thought to invoke high-level motion-processing mechanisms, it is likely that high-level motion processing occurs in at least these areas.

The LOC has been implicated in the processing of form (Malach et al., 1995; Kanwisher et al., 1996; Grill-Spector et al., 2001; Haxby et al., 2001). Evidence is emerging that LOC processes global 3D object shape, rather than local 2D shape features (Grill-Spector et al., 1998, 1999; Malach et al., 1998; Mendola et al., 1999; Gilaie-Dotan et al., 2001; Moore and Engel, 2001; Avidan et al.,

2002; Kourtzi and Kanwisher, 2000b, 2001; Kourtzi et al., 2003), and may even mediate aspects of object recognition (Grill-Spector et al., 2000). Indeed TAM itself occurs over 3D representations of form (Tse and Logothetis, 2002). The fMRI data concerning suggest that the LOC plays a greater role in TAM than the control. While both the TAM condition and the control condition involve global forms, the TAM condition may place a greater workload upon the LOC because the global form or global figural relationships that are believed to be computed there are presumably fed into motion-processing areas in the TAM case, whereas this output is not required for the control case. While fMRI data cannot specify the temporal dynamics of interactions among these areas, a reasonable model would place contour-based form analysis in the LOC and in retinotopic areas. The results of parsing would then be sent to hMT+ where motion trajectories would be computed in light of both figural matching and motion-energy cues.

These findings contrast with recent fMRI findings that used translational apparent motion as a probe. Liu et al. (2004) found no difference between apparent motion and flicker conditions in any retinotopic area, although they did see greater activation for apparent motion than flicker in hMT+. They conclude from this that there is no evidence for the filling-in of features along the path of perceived motion in early retinotopic areas. Other fMRI studies have found greater activation for apparent motion than flicker in hMT+, but not in V1 (Goebel et al., 1998; Muckli et al., 2002). While both hMT+ and V1 contain motion sensitive cells, the types of motion to which these cells respond need not be the same, perhaps accounting for this difference. For example, Mikami et al. (1986) found that neurons in MT but not in V1 responded to long-range apparent motion. Since the present data reveal differences between TAM and flicker in early retinotopic areas, it is possible that TAM, unlike translational apparent motion, does involve filling-in of features in early retinotopic areas.

These findings suggest that hMT+ should be thought of as part of a form/motion-processing stream of analysis, rather than as an area just dedicated to the processing of motion, and in

particular, just motion-energy. Similarly, these findings suggest that the LOC should be thought of as part of a form/motion-processing stream of analysis, rather than as an area just dedicated to form processing. Indeed, a number of recent papers have come to the conclusion that there is potential anatomical and functional overlap between hMT+ and the LOC (Ferber et al., 2003; Kourtzi et al., 2003; Murray et al., 2003; Zhuo et al., 2003; Liu et al., 2004) supported by behavioral data as well (Stone, 1999; Liu and Cooper, 2003). Because the form processing that underlies TAM must be extremely rapid, it would appear that LOC, hMT+, and other areas involved in form/motion processing operate in conjunction to solve the problem of what went where in TAM before TAM is perceived.

The unification of hMT+ and LOC into interacting motion-from-form analyzers blurs the distinction between the “what” vs. “where” pathways. Traditionally it has been maintained that form processing is a ventral or “what” stream process (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992), whereas motion processing is a dorsal or “where/whence” stream process. While the ventral/dorsal distinction is a useful heuristic for understanding gross-scale information-processing architecture, the current data imply that ventral and dorsal processing are not independent, and interact in a bottom-up manner in the few hundred milliseconds between sensory input and the conscious perception of TAM.

Implications for models of TAM

Baloch and Grossberg (1997) modeled TAM to involve three interacting subprocesses: (1) a boundary completion process (V1 → interstripe V2 → V4); (2) a surface filling-in process (blob V1 → thin stripe V2 → V4); and (3) a long-range apparent motion process (V1 → MT → MST), adding a link between V2 and MT in their model to allow the motion-processing stream to track emerging boundaries and filled-in surface colors. The model is foremost a bottom-up account of boundary completion, where existing boundaries enhance growth of neighboring collinear boundaries and

inhibit growth of neighboring dissimilarly oriented boundaries. Color is then filled in from existing color regions to new color boundaries. The data from our recent fMRI experiments cannot distinguish BOLD responses in V2 due to thin stripe vs. thick stripe activity, and can therefore not precisely test the model. However, the present fMRI data implicate all the areas forecast by this model. In addition to areas of activation predicted by this theory, V3 V, V3A/B, and LOC were found to be more responsive to TAM than the control, suggesting that the model is incorrect, or at best incomplete. Another difficulty for the model emerges from the local nature of the boundary interactions they describe. TAM is influenced by global configural relationships among stimuli. For example, in the stimuli in our fMRI experiments (Fig. 2), the direction of perceived motion in the central bar is influenced by the location of the square many degrees away. This means that the mechanism underlying TAM must be sensitive to global configural relationships within and between the two images. Purely local lateral excitation and inhibition are not sufficient to account for TAM (this applies also to Zanker, 1997).

TAM highlights a central problem in visual neuroscience: how does local feature information become integrated into a global representation of spatiotemporal figural relationships, which in turn appears to influence how local features are interpreted? TAM should prove a useful probe in future studies that attempt to answer this difficult question first raised by the Gestalt psychologists nearly a century ago (e.g., Wertheimer, 1912; Kenkel, 1913). TAM requires contour integration over large scales and a global analysis of contour relationships in order to determine the correct correspondence of figures over time, such that TAM can take place over corresponding figures (see Fig. 1). fMRI studies that have examined how contours are integrated into global shapes reveal that contour integration activates V1 and V2 in humans and monkeys but produces strongest activation in lateral occipital areas (Altmann et al., 2003; Kourtzi et al., 2003). That striate cortex shows BOLD activation for a task that requires global contour integration is consistent with recent work in neurophysiology. However, this is perhaps surprising, given that to date

V2 is the first area in the visual hierarchy where illusory contours have been shown to have a direct effect using single-unit recording (von der Heydt et al., 1984). While processing in V1 has traditionally (Hubel and Wiesel, 1968) been thought to be limited to the processing of local features, recent evidence has implicated early visual areas such as V1 and V2 in the processing of global shape (Allman et al., 1985; Gilbert, 1992, 1998; see also Fitzpatrick, 2000; Lamme et al., 1998, for reviews). In light of these findings, it is possible that the global form processing that underlies TAM begins as early as V1. The recent fMRI data examining the neural basis of TAM (Tse, 2006) indeed does find greater V1 activation to TAM than a no-TAM control. However, this cannot be used to establish, for example, that global processing takes place in V1. These data are also not sufficient to determine whether the activity seen in V1 arises because of bottom-up or top-down activation. The BOLD activity seen is consistent with both possibilities and indeed may arise from both mechanisms.

It is widely recognized that Fourier decomposition of motion-energy and pooling of motion signals at a single location are not capable of solving all motion-processing problems (e.g., Born and Bradley, 2005). More than one moving object can appear in a single region in space, and, on the retina, motions from multiple depth planes can project to the same retinal location, as occurs in cases of transparency. Pooling motions across depths or objects would create spurious motion signals, since the true motions can be independent. Thus, parsing mechanisms must exist that segregate motion signals from distinct objects or depth layers from each other for separate motion trajectory computations (Hildreth et al., 1995; Nowlan and Sejnowski, 1995). One possibility is that the parsing takes place within MT (hMT+) itself. There is evidence that motion-processing cells may contribute to such parsing procedures in that MT cells are highly tuned to depth and the response suppression that usually occurs with nonpreferred motion is attenuated when nonpreferred motion lies on a different depth plane from the preferred motion (Bradley et al., 1995). Moreover, whether overlapping sine waves are manipulated to look like a plaid or two overlapping transparent layers

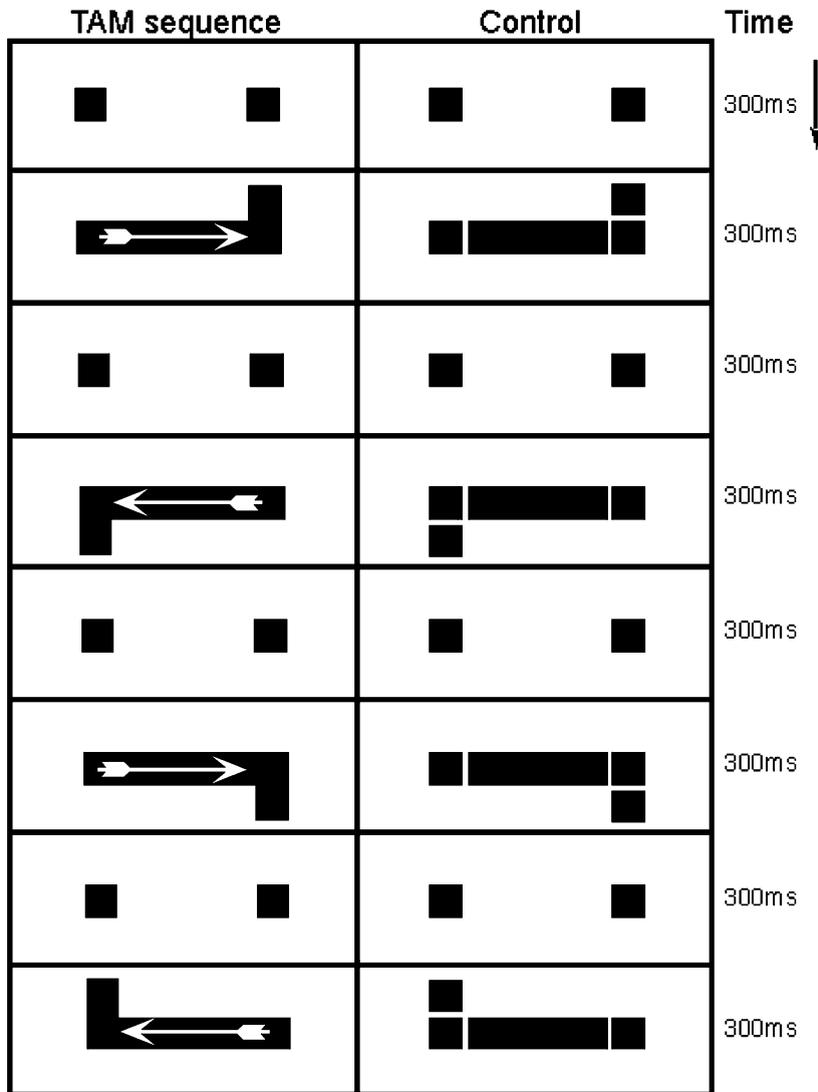


Fig. 2. Two conditions were tested, one where TAM was seen, and the other, where no TAM or other motion was seen. In both cases, the stimuli consisted of identically sized white squares ($1.68^\circ \times 1.68^\circ$) and bars ($5.40^\circ \times 1.68^\circ$) on a black background. In the TAM case, the bars and squares abutted. In the control case they were separated by a small gap (0.32°). The direction of perceived motion is indicated with black arrows (no arrows were actually shown to observers). A small fixation point ($0.2^\circ \times 0.2^\circ$, not shown) was located just below (0.30°) the lower contour of the bar.

by changing the luminance of x -junctions effects whether individual MT cells respond with pattern- or component-like directional tuning (Stoner and Albright, 1992, 1996). The typical motion opponency observed in MT cells may play a role in segregating figures moving relative to a moving background. However, the present fMRI data imply that it is more likely that at least some aspects

of parsing, particularly those involved in segmenting figures on the basis of global contour and other form relationships, do not take place in hMT+. It is more likely that areas specialized in such form processing, such as the LOC, carry out this type of global figural segmentation and then constrain motion processing in hMT+ to be consistent with figural identity matches over time.

Finally, the steps of parsing figures in space and matching figures in time are not separable. While it is conceivable that form analysis would operate on each image independently, followed by an independent stage of matching based on form correspondence among elements in successive images, form analysis and correspondence matching are not independent or serial processes. Because the form interpretation reached for a given figure takes into account the form of figures in the previous scene, form analysis and matching are part and parcel of the same spatiotemporal process. Matching occurs because figures are defined spatiotemporally, rather than just spatially in each successive scene. In other words, matching is subsumed by a figure formation process that operates over a certain range of spatiotemporal extents, and what gets matched across scenes is the corresponding 3D figures. Thus, form as processed by the visual system, is spatiotemporal rather than just spatial. Therefore, the existing dichotomy between (spatial) form or figure analysis and the (temporal) matching of figures across images is a misleading one. The first proponent of an entirely spatiotemporal analysis of form and motion may have been Gibson (1979). Recent work emphasizing the importance of spatiotemporal form and motion processing can be found in Gepshtein and Kubovy (2000) and Wallis and Bülthoff (2001).

Contour discontinuities as trackable features

So far we have been emphasizing that contour cues, in particular cues associated with contour continuity and discontinuity, are essential to figural segmentation, matching, and tracking, and that this stage of form analysis is central to high-level motion processing. In this section we will discuss how contour cues such as regions of high curvature and curvature discontinuities can serve as trackable features (TFs), disambiguating the aperture problem as it arises in the processing of rotational motion.

Motion perception is beset with the problem that motion signals are inherently ambiguous. There are many possible motions in the world that could have given rise to any particular motion that could be measured at the level of the retina or later. A key

cause of this ambiguity is the so-called “aperture problem,” which arises because receptive fields in early stages of the visual processing hierarchy are small. Because they are small, they can only recover the component of motion perpendicular to the orientation of a contour. At the heart of the problem is the fact that an infinite number of 3D velocity fields can generate the same 2D retinal sequence. The local motion information at any point along a contour is consistent with an infinite number of possible motions that all lie on a “constraint line” in velocity space (Adelson and Movshon, 1982) for the 2D case. The problem of interpreting this many-to-one mapping is commonly termed the “aperture problem” (Fennema and Thompson, 1979; Adelson and Movshon, 1982; Marr, 1982; Nakayama and Silverman, 1988).

Explaining how the aperture problem is solved is perhaps the most basic challenge that must be met by any model of motion perception. While there are several theoretical solutions to the aperture problem that account for many aspects of motion perception, no single general theory has yet emerged that can explain how the visual system actually processes motion in every instance. Many authors have argued that the aperture problem can be solved by integrating component motion signals along the contour (Bonnet, 1981; Burt and Sperling, 1981; Adelson and Movshon, 1982; Watson and Ahumada, 1985). These models are based on the assertion that ambiguous motion signals can, via integration, be disambiguated. Several models that provide reasonable solutions to the aperture problem for the case of translational motion, such as “intersection of constraints” and “vector summation” models, fail to provide unique solutions in the case of rotational motion.

An account that does provide a solution to the aperture problem in the case of both translational and rotational motion is that based upon the tracking of TFs (Ullman, 1979). TFs disambiguate ambiguous component motion signals that arise along portions of contour distant from TFs (Ullman, 1979) because these locations along a contour such as corners, terminators, and junctions do not move ambiguously when they are intrinsically part of the moving object (Shimojo et al., 1989; i.e., terminators that arise from endpoints

belonging to the moving line vs. those arising from points where the moving line is occluded). Recent work has shown how such terminator motions influence processes such as amodal completion and global integration of local motion signals (Lorenceau and Shiffrar, 1992; Shiffrar et al., 1995).

Relying solely on such form features, however, creates new problems that are in some ways as problematic as the aperture problem they are meant to solve. For one, corners, terminators, and junctions can arise in the image for reasons of occlusion that have nothing to do with the motion of the stimulus. These “extrinsic” form features give rise to spurious motion signals that can lead to incorrect conclusions about what motion is actually taking place in the world (Shimojo et al., 1989). The human brain appears to get around the ambiguities of motion stimuli by having multiple motion systems that process different characteristics of the moving stimulus, namely,

one based upon motion-energy processing and another that processes form cues.

Contour curvature is a vital cue for the analysis of both form and motion. We recently conducted a series of fMRI experiments (Caplovitz and Tse, 2006) designed to isolate the neural circuitry underlying the processing of CC as a TF for the perception of motion in general and rotational motion in particular. Using stimuli, called bumps (Tse and Albert, 1998; Kristjansson and Tse, 2001), in each of the experiments the degree of CC was parametrically modulated across stimulus group, as shown in Fig. 3. Each of the three experiments controlled for a different aspect of the visual stimulus, and together isolated CC as the sole form cue, which varied across stimulus condition. A region of interest analysis was applied to the data from each experiment identify brain areas in which the BOLD signal was systematically modulated as a function of CC. We found that the

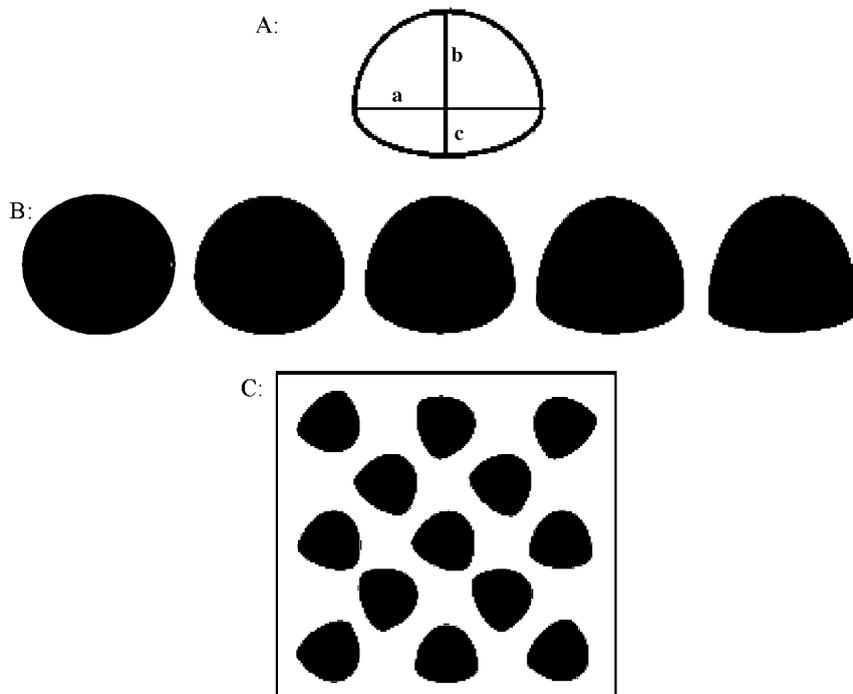


Fig. 3. (A) The bump stimulus is constructed by combining two half-ellipses along their common major axis ‘a.’ The degree of contour curvature and severity of contour curvature discontinuities can be controlled for by carefully selecting the relative sizes of the two minor axes ‘b’ and ‘c’ while still preserving the overall area of the stimulus. (B) In each of the fMRI experiments contour curvature was parametrically varied across each of the five stimulus groups shown here. (C) Stimuli as presented in the scanner: in each stimulus block, 13 rotating bumps were presented continually for 20 s.

BOLD signal in retinotopic area V3A varied parametrically with the degree of CC. The convergent results of the three fMRI experiments ruled out the possibility that these modulations resulted from changes in the area of the stimuli, the velocity with which contour elements were translating, and perceived angular velocity. We concluded that neurons within V3A process continuously moving CC as a TF, and this form-based feature is used to determine the speed and direction of motion of the objects as a whole. In this context, it is suggested that V3A contains neural populations that process form, not to solve the “ventral problem” of determining object shape, but in order to solve the “dorsal problem” of what is going where.

V3A is an area that several studies have found to be motion selective (Tootell et al., 1997; Vanduffel et al., 2002). Schira et al. (2004) demonstrated that %BOLD signal in V3A is also correlated with contour and figural processing, even in the absence of conscious perception. Figural processing is central to the TF argument, as the motion signal derived from the TF must be generalized to the rest of the contour.

V3A has also been identified as a neural correlate of form–motion interactions. Several groups (Braddick et al., 2000, 2001; Vaina et al., 2003; Moutoussis et al., 2005) have shown that %BOLD signal change in V3A was greater for coherent than for random motion. Koyama et al. (2005) showed V3A is more responsive to radial than for translational motion. These findings suggest a role for V3A in the generation of global motion percepts. It is notable that the fMRI investigation of TAM described earlier in this chapter found, in addition to other areas, V3A to be more active during the TAM percept relative to control.

The recent finding that BOLD signal in area V3A is modulated by CC in the context of rotational motion expanded upon this previous work by suggesting a specific mechanism concerning how form and motion may interact to construct global motion percepts. Namely, we hypothesize that neural activity within V3A serves to extract reliable motion information from regions of high CC. Such TF motion information may then be propagated to the entire moving object, resulting in the global motion percept.

How might TFs be processed? Recent neurophysiological data have shown that neurons in MT in the macaque respond more to terminator motion in a barber pole stimulus than to the ambiguous signals generated by contours. Furthermore, they respond more to intrinsically owned terminators than to extrinsic terminators (Pack et al., 2004). It has also been shown that neurons in MT in the macaque will initially respond to the direction of motion that is perpendicular (component direction) to a moving line independent of the actual direction of motion (Pack and Born 2001). These same neurons will, over a period of approximately 60 ms, shift their response properties so that they respond to the true motion of the line independent of its orientation, suggesting that the unambiguously moving endpoints of the line are quickly but not instantaneously exploited to generate a veridical motion solution. The response properties of these neurons match behavioral data that show initial pursuit eye movements will be in the direction perpendicular to the moving line, and then rapidly adapt to follow the direction of veridical motion as defined by the terminators of the lines (Pack and Born, 2001). There is also neurophysiological evidence of end-stopped neurons in V1 that respond to the motion of line terminators independently of the line’s orientation (Pack et al., 2003), suggesting that form-based TFs such as line terminators can be directly extracted from the image as early as V1. Such cells are largely immune to the aperture problem.

In line with this view, features to which such end-stopped cells would respond have been shown psychophysically to be processed both rapidly and in parallel across the visual scene. Visual search studies have found several form-based features, including certain types of contour junctions (Enns and Rensink, 1991), contour concavities (Hulleman et al., 2000), corners (Humphreys et al., 1994), CCs (Wolfe et al., 1992) and curvature discontinuities (Kristjansson and Tse, 2001), which will pop out among a set of distractors. It is commonly believed that features that exhibit pop out during visual search are processed rapidly and in parallel across the visual field (Treisman and Gelade, 1980), suggesting the existence of hard-wired contour discontinuity detectors in V1 or

later. Indeed, contour discontinuity information may be extracted even before V1, since circular center-surround receptive fields will respond more to corners than to edges, and more to bar terminators than corners (Troncoso et al., 2005).

However, simply because local features pop out in visual search experiments does not mean that they pop out because there is a dedicated local detector for this feature that is uninfluenced by global context. For example, relative maxima of positive curvature are inherently relational in character. They do not count as maxima because of their local curvature but because of their curvature relative to other CCs in a global figure. Furthermore, such features may arise due to occlusion, and the determination of whether a feature is intrinsic to the moving object or not requires non-local processes. Only a global analysis of form can adequately specify which contour discontinuities are intrinsically owned by a moving object and which are extrinsic, and therefore potential sources of spurious motion signal (Lee et al., 1998). Thus, even if end-stopped hypercomplex cells in V1 play an essential role in detecting CC, as first suggested by Hubel and Wiesel (1965; Dobbins et al., 1987), they cannot directly or locally isolate these contour elements as TFs of interest in moving figures. Similarly, the form features that underlie figural segmentation are inherently global and relational in character. It therefore appears that a more global analysis of form than that available in V1 or hMT+ must interact with the motion-processing system to account for TAM and to account for the extraction and analysis of intrinsically owned TFs.

Implications for the architecture of motion processing

The data from the experiments described shed light on a debate concerning the basic architecture of the motion-processing system that has concerned the field for more than 30 years. Early studies implied that there are two motion-processing subsystems, one called the “short-range” system and the other called the “long-range” system (Braddick, 1974). The short-range process was initially touted to operate over short distances and durations and

emerge from the responses of low-level motion-energy detectors that respond automatically, passively, and in parallel across the whole visual field, while the long-range process was described as operating over long distances and durations and was thought to emerge from the responses of a cognitive mechanism that identifies forms and then tracks them over time (Anstis, 1980; Braddick, 1980). Subsequent research showed that the presumed distinctions between these processing systems did not hold. Contrary to the initial delineation of the two subsystems, short-range motion could occur over large distances and color stimuli, and long-range motion over short distances (Cavanagh et al., 1985). Contrary to predictions, it was found that long-range motion can produce motion aftereffects (von Grünau, 1986).

The reason the short- vs. long-range motion-processing distinction foundered was that it confounded differences in processing type with differences in stimulus type. Cavanagh and Mather (1989) tried to remedy this confusion by arguing that while there was only one type of motion processing, namely, the detection of motion-energy, there were two types of motion stimuli, first order (luminance and color defined) and second order (texture, relative motion, and disparity defined). There were first-order motion-energy detectors that detected first-order stimuli and second-order motion-energy detectors that detected second-order stimuli.

This argument belonged to a tradition that is still alive today that attempts to reduce all of motion perception, ultimately, to the detection of motion-energy (Lelkens and Koenderink, 1984; Chubb and Sperling, 1988; Cavanagh and Mather, 1989; Johnston et al., 1992; Wilson et al., 1992). When certain stimuli cannot be detected by luminance-defined (first-order) motion-energy detectors, a “front end” of neuronal circuitry is posited that converts the signal into a change that a more or less straightforward motion-energy detector sitting on top of the front-end circuitry can detect, whether using a front end that carries out some form of rectification (as, supposedly, occurs in second-order motion processing; Chubb and Sperling, 1988; Solomon and Sperling, 1994), or salience mapping (as is posited to occur in

third-order motion processing; Lu and Sperling, 1995, 2001). Note that Sperling and colleagues have taken Cavanagh and Mather's (1989) initial distinction between first-order and second-order stimulus characteristics, and turned them into first-, second-, and third-order motion-processing systems, complicating the debate because terms like "second-order motion" are now used to mean different things by different authors.

Several authors (Mather and West, 1993; Ledgeway and Smith, 1994, 1995; Seiffert and Cavanagh, 1998) have provided evidence that first- and second-order motion are not processed by a common motion detector type. Rather, they are processed by separate low-level detectors, each insensitive to motion of the other class. Edwards and Badcock (1995), however, provided evidence that unlike first-order motion detectors, second-order motion detectors were sensitive to motion of the other class. This would be expected if second-order filters detect or are influenced by contour, texture, or form information, because form can be defined using first-order cues. Wilson et al. (1992; see also Derrington et al., 1993; Zanker and Huepfgens, 1994) suggested that first-order motion is processed foremost in V1, whereas second-order motion also requires processing in V2, with both streams converging in MT. This was supported by the finding that 87% of sampled cells in MT in the alert macaque are tuned to second-order motion as well as first-order motion (Albright, 1992, see also Olavarria et al., 1992; but compare O'Keefe and Movshon, 1998 who found <25% such cells in MT in the anaesthetized macaque; for second-order motion tuning in cat cortex see Mareschal and Baker, 1998, 1999; Zhou and Baker, 1993, 1994, 1996). Albright (1992) suggested that these cells may underlie form-cue invariant motion processing (see also, Buracas and Albright, 1996). Form-cue invariant TAM, and TAM more generally, may be processed in MT (cf. Kawamoto et al., 1997) by such cells, and may therefore have much in common with second-order motion. Some authors have reported areas of the brain that respond more to second-order motion than first-order motion. Smith et al. (1998) reported that V3 was more activated by second-order motion, and Dumoulin et al. (2003) found a region

posterior to hMT+ that was preferentially activated by second-order motion. Other authors (Seiffert et al., 2003) have found no difference in retinotopic areas, hMT+, or LOC in the processing of first- vs. second-order motion.

Several authors have argued that second-order motion processing requires a stage of texture extraction (Derrington and Henning, 1993; Stoner and Albright, 1993; Werkhoven et al., 1993). This stage of texture extraction (presumably carried out in V2, V3, and/or V4; compare Smith et al., 1998) may play a role in the stage of form analysis that must precede the assignment of motion paths in TAM. Of course, there need not be a single form processor underlying TAM just as there need not be a single stage of form extraction underlying second-order motion. There may exist several shape-from-*x* systems, just as there may be multiple types of second-order motion detectors (Petersik, 1995). But for form-cue invariant TAM to be possible (e.g., between a luminance-defined square and color-defined rectangle; Tse and Logothetis, 2002), the various shape-from-*x* systems and/or second-order motion detectors must converge on a common representation of moving shape. Texture features, once extracted, could play a role in the formation of the figural representations that are transformed in TAM.

Reducing all types of motion processing, ultimately, to the processing of motion-energy fails to account for the importance of spatiotemporal segmentation and tracking of figures in motion processing. TAM reveals a role of figural tracking in motion perception that cannot be readily reduced to motion-energy accounts. There are cases where TAM can be perceived in stimulus configurations for which there is no known motion-energy detector. For example, TAM is seen when replacing a square section of a random dot field with new random dots, and then replacing this square with a rectangle of new random dots 200 ms later (Hsieh and Tse, 2006). On each frame there is nothing but random dots, so no figure seems to exist when each frame is considered in isolation. The square and rectangular figures emerge in the relationship among successive random dot fields, and must involve a parser that can extract these figures on the basis of inferred

contour or surface information. This and other examples of TAM imply that there is a type of motion processing that cannot be reduced to low-level processing of local motion-energy.

TAM requires that a stage of figural segmentation and matching on the basis of global form cues exists that can determine the perceived direction of motion, even when motion-energy would predict the opposite direction of motion. For example, in Fig. 4, all motion-energy accounts would predict that motion should be perceived to the left, whereas TAM is perceived to proceed to the right, as indicated by the white arrow. Although there are many different motion-energy models, all have in common that the centroid of a luminance blob remains the centroid regardless of spatial frequency.

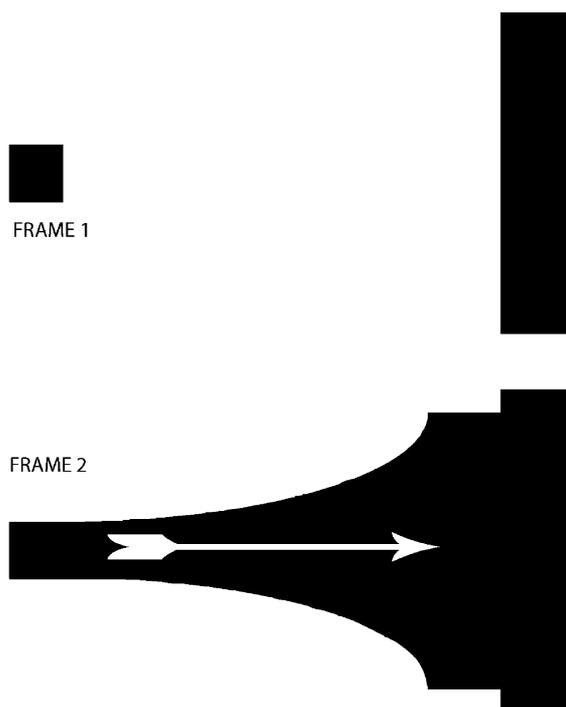


Fig. 4. Even though the center of the new image data in frame 2 is closer to the right-hand cue, motion proceeds away from the left-hand cue, implying that more is involved than the matching of luminance blobs, as is commonly assumed in motion-energy models. In particular, perceived TAM proceeds from left to right in this example because the new image data are matched as comprising a shape change in the left-hand cue, based on the location of continuous contours and deep concavities, which permit figural parsing and matching.

Moreover, all models of motion-energy agree that motion-energy is stronger from one centroid in the first frame to the nearest centroid in the subsequent frame (The “nearest neighbor principle”; Ullman, 1979). Centroid-to-centroid matching would imply motion to the left, whereas inferred motion that matches on the basis of figural identity across images would imply motion to the right. Because this type of high-level motion processing can violate a basic premise of all motion-energy models, it can be concluded that there are at least two mutually irreducible types of motion processing, one built upon the detection of local motion-energy (low-level, passive, parallel, monocular, automatic) and the other built upon the tracking (parsing and matching) of global figures based upon form relationships within and between successive images. A second type of form analysis that subserves motion processing is the analysis of TFs.

Reducing all types of motion processing, ultimately, to the processing of motion-energy also fails to take into account the importance of TFs in disambiguating low-level motion signals. It is likely that the paramount TFs used by the visual system to overcome the aperture problem are contour discontinuities, including terminators, junctions, corners, and curvature. These are manifestly not a motion-energy cue and cannot be reduced to motion-energy. Rather, like the stage of contour-based segmentation that underlies TAM, which apparently involves retinotopic areas and the LOC, the stage of TF analysis that we argue takes place in V3A must proceed in parallel with a motion-energy analysis. The neural computations that underlie the conscious experience of motion, which presumably take place in hMT+, combine inputs from the low-level, motion-energy-based system, and the two form-based processes described here. It is as if there is a ventral process of global contour-based form segmentation that culminates in the LOC, and a dorsal process that tracks key contour discontinuities that culminates in V3A. Both these form analyses apparently feed into hMT+ where the final motion trajectory is thought to be computed in light of the constraints that they provide.

Although Cavanagh (Cavanagh and Mather, 1989) initially argued that all motion processing

could be reduced to motion-energy processing, [Cavanagh \(1992\)](#) subsequently reported the existence of a new attention-based or high-level motion process that could not be reduced to an account in terms of motion-energy. [Cavanagh \(1995\)](#) has called the low-level and high-level systems the “passive” (low-level, motion-energy based for color, texture, and luminance) and “active” (based upon attentional tracking) motion-processing systems. Since TAM cannot be accounted for by the responses of the passive system, this raises the question of whether TAM requires not just form-based parsing, but also attention in order to be seen. It has been shown that TAM does not arise because of attentional gradients ([Hsieh et al., 2005](#)). However, other evidence is emerging that TAM requires the specification and tracking of a figure over time. It could be that TAM does indeed require attention in order to be perceived, not in the sense of attentional gradients, but in the sense that attention is required for the tracking of figural changes over successive scenes. In essence, attention converts what would otherwise be a mere succession of disjoint shapes ([Horowitz and Wolfe, 1998](#)) into a bound object representation that can undergo changes in location, shape, and other features, and that can be tracked over time ([Kahneman et al., 1992](#)).

To test whether attention is needed for the perception of TAM in the sense of figural tracking, a simple psychophysical experiment was carried out on TAM in the context of a standard change blindness paradigm. An array of squares and rectangles was continuously flashed on and off. On alternating flashes one of the squares would be replaced by a rectangle, which in turn would be replaced by the square on the next flash, and so on. While maintaining fixation, observers were asked to localize the figure undergoing shape changes. This is slow and requires a serial search of the whole array, as has been shown previously ([Rensink et al., 1997](#); [Rensink, 2000](#); [Tse et al., 2003](#)). However, if the location is cued, the change is perceived immediately even when the figure is located as far as fifteen degrees from fixation, in line with past findings ([Rensink et al., 1997](#)) and the general notion of attentional “pop-out.” What is new is that as soon as a change is seen, the figure

is seen to undergo TAM from a square into a rectangle or vice versa. This implies that TAM indeed does require attention because attention is required to bind successive shape representations into the representation of a single spatiotemporal figure that can undergo a shape change.

What does this tell us about motion processing? It tells us that the high-level motion-processing stream is one that tracks figural changes on the basis of changes in position or shape of various features, and it tells us that such figural tracking requires attention to bind successive disjoint ([Horowitz and Wolfe, 1998](#)) shape representations into a representation of a single object ([Kahneman et al., 1992](#)) undergoing a shape change. Thus the high-level motion-processing stream is inherently driven by form analysis and attentional tracking, and the low-level motion-processing stream is driven by the analysis of various classes of motion-energy. This is not to say that low-level motion processing is stimulus driven whereas high-level processing is not. Once attention has marked an object as a figure to be tracked, the fact that TAM can be processed so rapidly implies that form-based processes parse, match, and track figures over time as they change shape and position in a largely automatic and stimulus-driven manner.

How does the low-level/high-level dichotomy advocated here mesh with past views? TAM vindicates the initial speculations of [Braddick \(1980\)](#) and [Anstis \(1980\)](#) that the essence of the high-level motion-processing system involves matching on the basis of form. During the 1980s researchers found little evidence to support this view of high-level motion processing, and many came to the conclusion that all motion processing could be reduced to the processing of motion-energy detectors of various types, none of which were sensitive to form relationships per se. The present view is consistent with [Cavanagh’s](#) post-1992 view ([Cavanagh, 1992, 1995](#)) of passive and active motion-processing systems. The high-level motion-processing system described here is the same as his active vision system, as long as it is understood that the motion signal is not necessarily or primarily driven by the top-down movements of an attentional window, but rather by the outputs of a stimulus-driven analysis of form. Once a figure is

defined by attention (i.e., by its binding into an object representation; Kahneman et al., 1992), its tracking becomes largely stimulus driven.

The present view of motion processing is not entirely consistent with the hierarchy of motion-energy types posited by Sperling and colleagues (e.g., Lu and Sperling, 1995) because there is no place in their models for matching based upon the explicit processing of global form relationships, or for TFs. The saliency mapping posited to underlie third-order motion processing resembles attention, but nonetheless feeds into a more or less standard motion-energy detector. Even though the motion-energy detector detects shifts in the peaks of salience, such a detector would still map peaks in salience to their nearest neighbor. Such a detector would presumably predict leftward motion for the stimulus in Fig. 4. But because rightward motion is what observers actually perceive, matching must take place on the basis of something other than peak to peak or centroid-to-centroid matching, whether of luminance or of salience.

Lu and Sperling (2001) justify their claim that third-order motion involves motion-energy detection over shifts in salience on the basis of an evolutionary argument:

An important reason for assuming that third-order motion is computed by an algorithm similar to that for first- and second-order motion is that the genetic code needed to instantiate a computation in the brain is quite complex. The likelihood that a new gene for a motion computation would evolve separately vs. the original motion computation gene being spliced from one location to another is negligibly small. (p. 2335)

In contrast, a central claim of the present paper is that the figural tracking underlying high-order motion processing is not based upon motion-energy detection at all, but is based upon a very different algorithm premised upon spatiotemporal figural parsing, matching, and binding on the basis of global form cues.

In addition, it may not be true that only third-order motion is salience- or attention based, whereas second-order motion is not. While Lu and

Sperling (1995) argue that neither first- nor second-order motion depends on tracking, there is evidence that at low contrasts, second-order motion is dependent on figural tracking (Seiffert and Cavanagh, 1998; Derrington and Ukkonen, 1999). In addition, it is likely that some kinds of stimuli classified as second-order in the past have unwittingly relied on the type of form analysis that underlies TAM. Thus the trichotomy into first-, second-, and third-order motion-processing systems may complicate matters by viewing all three systems as involving motion-energy detection, by segregating attentional mechanisms from second-order mechanisms, and by ignoring the contribution of figural analysis based on global form relationships to motion processing. Any model of motion processing that fails to explicitly incorporate matching on the basis of global form-relationships within and between successive scenes will be incomplete. A simpler architecture may be adequate to explain all motion data, where a dichotomy exists between a low-level system which processes motion in terms of motion-energy derived from various features such as luminance, color, and texture, and a high-level system that processes motion in terms of changes over figures that are parsed and matched on the basis of their global form relationships rather than salience.

Conclusion

Contour discontinuities, such as corners, deep concavities, maxima of positive curvature, junctions, and terminators, are the most informative portions of an object's contour used in determining perceived object shape (e.g., Tse, 2001) and motion (e.g., Tse, 2006, Caplovitz and Tse, 2006; Caplovitz, Hsieh and Tse, 2006). Although contour discontinuities are a form cue, they play at least two important roles in determining the direction of perceived motion. First, they play a central role in a stage of figural segmentation and matching that must precede the perception of motion because it can determine the direction of perceived motion. TAM was discussed in this context. This is primarily a ventral process involving the LOC and probably also V2 and V4, and perhaps

even V1. Second, contour discontinuities play an important role as TFs that can solve the aperture problem. Their role in the perception of rotating figures was considered. This TF analysis is a dorsal process involving at least V3A (Caplovitz and Tse, submitted). In this context, it is suggested that area V3A contains neural populations that process form, not to solve the “ventral problem” of determining object shape, but in order to solve the “dorsal problem” of what is going where. To put it more succinctly, form is involved in motion processing at least two ways, in terms of figural segmentation dedicated to solving the problem of figure to figure matching over time, and in terms of defining TFs whose unambiguous motion signals can be generalized to ambiguously moving portions of an object. In both cases, the analysis of form must proceed in parallel with the analysis of motion, in order to constrain the ongoing analysis of motion. The neural computations that underlie the conscious experience of motion, which presumably take place in hMT+, combine inputs from the low-level, motion-energy based system, and the two form-based processes described here. There is a ventral process of global contour-based form segmentation that culminates in the LOC, and a dorsal process that tracks key contour discontinuities that culminates in V3A. Both these form analyses apparently feed into hMT+ where the final motion trajectory is presumably computed in light of the constraints that they provide.

References

- Adelson, E.H. and Bergen, J.R. (1985) Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, 2(2): 284–299.
- Adelson, E.H. and Movshon, J.A. (1982) Phenomenal coherence of moving visual patterns. *Nature*, 30: 523–525.
- Albright, T.D. (1992) Form-cue invariant motion processing in primate visual cortex. *Science*, 255: 1141–1143.
- Allman, J.M., Miezin, F. and McCuiness, E. (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Ann. Rev. Neurosci.*, 8: 407–430.
- Altmann, C.F., Bühlhoff, H.H. and Kourtzi, Z. (2003) Perceptual organization of local elements into global shapes in the human visual cortex. *Curr. Biol.*, 13(4): 342–349.
- Anstis, S.M. (1980) The perception of apparent motion. *Philos. Trans. R. Soc. Lond. B*, 290: 153–168.
- Attneave, F. (1954) Some informational aspects of visual perception. *Psychol. Rev.*, 61: 183–193.
- Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E. and Malach, R. (2002) Contrast sensitivity in human visual areas and its relationship to object recognition. *J. Neurophysiol.*, 87: 3102–3116.
- Baloch, A.A. and Grossberg, S. (1997) A neural model of high-level motion processing: line motion and form-motion dynamics. *Vision Res.*, 37(21): 3037–3059.
- Barlow, H.B. and Levick, W.R. (1965) The mechanism of directionally selective units in rabbit’s retina. *J. Physiol.*, 178(3): 477–504.
- Baro, J.A. and Levinson, E. (1988) Apparent motion can be perceived between patterns with dissimilar spatial frequencies. *Vision Res.*, 28: 1311–1313.
- Biederman, I. (1987) Recognition-by-components: a theory of human image understanding. *Psychol. Rev.*, 94(2): 115–147.
- Bonnet, C. (1981) Processing configurations of visual motion. In: Long, J. and Bradley, A. (Eds.), *Attention and Performance IX*. Erlbaum, Hillsdale, NJ.
- Born, R.T. and Bradley, D.C. (2005) Structure and function of visual area MT. *Ann. Rev. Neurosci.*, 28: 157–189.
- Braddick, O. (1974) A short-range process in apparent motion. *Vision Res.*, 14(7): 519–527.
- Braddick, O. (1980) Low-level and high-level processes in apparent motion. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.*, 290(1038): 137–151.
- Braddick, O.J., O’Brien, J.M., Wattam-Bell, J., Atkinson, J. and Turner, R. (2000) Form and motion coherence activate independent but not dorsal/ventral segregated, networks in the human brain. *Curr. Biol.*, 10: 731–734.
- Braddick, O.J., O’Brien, J.M., Wattam-Bell, J., Atkinson, J., Hartley, T. and Turner, R. (2001) Brain areas sensitive to coherent visual motion. *Perception*, 30: 61–72.
- Bradley, D.C., Qian, N. and Andersen, R.A. (1995) Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature*, 373(6515): 609–611.
- Buracas, G.T. and Albright, T.D. (1996) Contribution of area MT to perception of three-dimensional shape: a computational study. *Vision Res.*, 36(6): 869–887.
- Burr, D.C., Ross, J. and Morrone, M.C. (1986) Seeing objects in motion. *Proc. R. Soc. Lond.*, B227: 171–195.
- Burt, P. and Sperling, G. (1981) Time, distance, and feature trade-offs in visual apparent motion. *Psychol. Rev.*, 88: 171–195.
- Caplovitz, G.P. and Tse, P.U. (2006) V3A processes contour curvature as a trackable feature for the perception of rotational motion. *Cereb. Cortex*, in press.
- Caplovitz, G.P., Hsieh, P.-J. and Tse, P.U. (2006) Mechanisms underlying the perceived angular velocity of a rigidly rotating object. *Vision Res.*, Apr. 27 [Epub ahead of print].
- Cavanagh, P. (1992) Attention-based motion perception. *Science*, 257: 1563–1565.
- Cavanagh, P. (1995). Is there low-level motion processing for non-luminance-based stimuli? In: Thomas, V. Papatomas, Charles Chubb, Andrei Gorea and Eileen Kowler (Eds.),

- Early Vision and Beyond. MIT Press, Cambridge, MA, pp. 113–120.
- Cavanagh, P., Arguin, M. and von Grünau, M. (1989) Inter-attribute apparent motion. *Vision Res.*, 29(9): 1197–1204.
- Cavanagh, P. and Mather, G. (1989) Motion: the long and short of it. *Spatial Vis.*, 4: 103–129.
- Cavanagh, P., Boeglin, J. and Favreau, O.E. (1985) Perception of motion in equiluminous kinematograms. *Perception*, 14(2): 151–162.
- Chubb, C. and Sperling, G. (1988) Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *J. Opt. Soc. Am. A*, 5(11): 1986–2007.
- Dawson, M.R.W. (1991) The how and why of what went where in apparent motion: modeling solutions to the motion correspondence problem. *Psychol. Rev.*, 33(4): 569–603.
- Derrington, A.M., Badcock, D.R. and Henning, G.B. (1993) Discriminating the direction of second-order motion at short stimulus durations. *Vision Res.*, 33(13): 1785–1794.
- Derrington, A.M. and Henning, G.B. (1993) Detecting and discriminating the direction of motion of luminance and colour gratings. *Vision Res.*, 33(5–6): 799–811.
- Derrington, A.M. and Ukkonen, O.I. (1999) Second-order motion discrimination by feature-tracking. *Vision Res.*, 39(8): 1465–1475.
- Dobbins, A., Zucker, S.W. and Cynader, M.S. (1987) End-stopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329(6138): 438–441.
- Dumoulin, S., Baker, C.L., Hess, R.F. and Evans, A.C. (2003) Cortical specialization for processing first- and second-order motion. *Cereb. Cortex*, 13(12): 1375–1385.
- Downing, P. and Treisman, A. (1995) The shooting line illusion: attention or apparent motion? *Invest. Ophthalmol. Vis. Sci.*, 36: S856.
- Downing, P. and Treisman, A. (1997) The line motion illusion: attention or impletion? *J. Exp. Psychol. Hum. Percept. Perform.*, 23(3): 768–779.
- Edwards, M. and Badcock, D.R. (1995) Global motion perception: no interaction between the first- and second-order motion pathways. *Vision Res.*, 35(18): 2589–2602.
- Enns, J.T. and Rensink, R.A. (1991) Preattentive recovery of three-dimensional orientation from line drawings. *Psychol. Rev.*, 98(3): 335–351.
- Faubert, J. and von Grünau, M. (1995) The influence of two spatially distinct primers and attribute priming on motion induction. *Vision Res.*, 35(22): 3119–3130.
- Fennema, C. and Thompson, W. (1979) Velocity determination in scenes containing several moving objects. *Comput. Graph. Image Process.*, 9: 301–305.
- Ferber, S., Humphrey, G.K. and Vilis, T. (2003) The lateral occipital complex subserves the perceptual persistence of motion-defined groupings. *Cereb. Cortex.*, 13: 716–721.
- Fitzpatrick, D. (2000) Seeing beyond the receptive field in primary visual cortex. *Curr. Opin. Neurobiol.*, 10: 438–443.
- Gepshtein, S. and Kubovy, M. (2000) The emergence of visual objects in spacetime. *Proc. Natl. Acad. Sci. U S A*, 97(14): 8186–8191.
- Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.
- Gilaie-Dotan, S., Ullman, S., Kushnir, T. and Malach, R. (2001) Shape-selective stereo processing in human object-related visual areas. *Hum. Brain Map.*, 15: 67–79.
- Gilbert, C.D. (1992) Horizontal integration and cortical dynamics. *Neuron*, 9: 1–13.
- Gilbert, C.D. (1998) Adult cortical dynamics. *Physiol. Rev.*, 78: 467–485.
- Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H. and Singer, W. (1998) The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur. J. Neurosci.*, 10: 1563–1573.
- Goodale, M. and Milner, A. (1992) Separate visual pathways for perception and action. *Trends Neurosci.*, 15: 20–25.
- Green, M. (1986a) Correspondence in apparent motion: defining the heuristics. *Proc. Vis. Interface*, 337–242.
- Green, M. (1986b) What determines correspondence strength in apparent motion? *Vision Res.*, 26: 599–607.
- Green, M. (1989) Color correspondence in apparent motion. *Percept. Psychophys.*, 45(1): 15–20.
- Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y. and Malach, R. (1998) Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, 21: 191–202.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y. and Malach, R. (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24: 187–203.
- Grill-Spector, K., Kushnir, T., Hendler, T. and Malach, R. (2000) The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.*, 3(8): 837–843.
- Grill-Spector, K., Kourtzi, Z. and Kanwisher, N. (2001) The lateral occipital complex and its role in object recognition. *Vision Res.*, 41: 1409–1422.
- Hassenstein, B. and Reichardt, W. (1956) Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenbewertung bei der bewegungsperzeption des russekafers *chlorophanus*. *Z. Naturforsch.*, 11b: 513–524.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L. and Pietrini, P. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539): 2425–2430.
- Hikosaka, O., Miyauchi, S. and Shimojo, S. (1991) Focal visual attention produces motion sensation in lines. *Investig. Ophthalmol. Vis. Sci.*, 32(4): 176.
- Hikosaka, O., Miyauchi, S. and Shimojo, S. (1993a) Focal visual attention produces illusory temporal order and motion sensation. *Vision Res.*, 33(9): 1219–1240.
- Hikosaka, O., Miyauchi, S. and Shimojo, S. (1993b) Visual attention revealed by an illusion of motion. *Neurosci. Res.*, 18(1): 11–18.
- Hildreth, E.C., Ando, H., Andersen, R.A. and Treue, S. (1995) Recovering three-dimensional structure from motion with surface reconstruction. *Vision Res.*, 35(1): 117–137.

- Horowitz, T.S. and Wolfe, J.M. (1998) Visual search has no memory. *Nature*, 394(6693): 575–577.
- Hsieh, P.-J., Caplovitz, G.P. and Tse, P.U. (2005) Illusory rebound motion and the motion continuity heuristic. *Vision Res.*, 45(23): 2972–2985.
- Hsieh, P.-J. and Tse, P.U. (2006) Stimulus factors affecting illusory rebound motion. *Vision Res.*, 46(12): 1924–1933.
- Hubel, D.H. and Wiesel, T.N. (1965) Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.*, 28: 229–289.
- Hubel, D.H. and Wiesel, T.N. (1968) Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.*, 195: 215–243.
- Hulleman, J., te Winkel, W. and Boselie, F. (2000) Concavities as basic features in visual search: evidence from search asymmetries. *Percept. Psychophys.*, 62(1): 162–174.
- Humphreys, G.W., Keulers, N. and Donnelly, N. (1994) Parallel visual coding in three dimensions. *Perception*, 23(4): 453–470.
- Johnston, A., McOwan, P.W. and Buxton, H. (1992) A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proc. Biol. Sci.*, 250(1329): 297–306.
- Kahneman, D., Treisman, A. and Gibbs, B.J. (1992) The reviewing of object files: Object-specific integration of information. *Cogn. Psychol.*, 24: 175–219.
- Kanizsa, G. (1951) Sulla polarizzazione del movimento gamma [The polarization of gamma movement]. *Arch. Psichol., Neurol. Psichiatr.*, 3: 224–267.
- Kanizsa, G. (1971) *Organization in Vision: Essays on Gestalt Perception*. Praeger, New York.
- Kanwisher, N., Chun, M.M., McDermott, J. and Ledden, P.J. (1996) Functional imaging of human visual recognition. *Brain Res. Cogn. Brain Res.*, 5(1–2): 55–67.
- Kenkel, F. (1913) Untersuchungen über den Zusammenhang zwischen Erscheinungsgröße und Erscheinungsbewegung bei einigen sogenannten optischen Täuschungen. *Z. Psychol.*, 67: 358–449.
- Kolers, P.A. and Pomerantz, J.R. (1971) Figural change in apparent motion. *J. Exp. Psychol.*, 87: 99–108.
- Kourtzi, Z. and Kanwisher, N. (2000b) Cortical regions involved in perceiving object shape. *J. Neurosci.*, 20: 3310–3318.
- Kourtzi, Z. and Kanwisher, N. (2001) Representation of perceived object shape by the human lateral occipital complex. *Science*, 293: 1506–1509.
- Kourtzi, Z., Erb, M., Grodd, W. and Bülthoff, H.H. (2003) Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cereb. Cortex*, 13(9): 911–920.
- Kourtzi, Z., Tolia, A.S., Altmann, C.F., Augath, M. and Logothetis, N.K. (2003) Integration of local features into global shapes. Monkey and human fMRI studies. *Neuron*, 37(2): 333–346.
- Koyama, S., Sasaki, Y., Andersen, G.J., Tootell, R.B., Matsuura, M. and Watanabe, T. (2005) Separate processing of different global-motion structures in visual cortex is revealed by fMRI. *Curr. Biol.*, 15(22): 2027–2032.
- Kristjansson, A. and Tse, P.U. (2001) Curvature discontinuities are cues for rapid shape analysis. *Percept. Psychophys.*, 63(3): 390–403.
- Lamme, V.A., Super, H. and Spekreijse, H. (1998) Feedforward, horizontal, and feedback processing in the visual cortex. *Curr. Opin. Neurobiol.*, 8: 529–535.
- Ledgeway, T. and Smith, A.T. (1994) Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision. *Vision Res.*, 34(20): 2727–2740.
- Ledgeway, T. and Smith, A.T. (1995) The perceived speed of second-order motion and its dependence on stimulus contrast. *Vision Res.*, 35(10): 1421–1434.
- Lee, T.S., Mumford, D., Romero, R. and Lamme, V.A. (1998) The role of the primary visual cortex in higher level vision. *Vision Res.*, 38(15–16): 2429–2454.
- Lelkens, A.M. and Koenderink, J.J. (1984) Illusory motion in visual displays. *Vision Res.*, 24(9): 1083–1090.
- Liu, T. and Cooper, L.A. (2003) Explicit and implicit memory for rotating objects. *J. Exp. Psychol. Learn. Mem. Cogn.*, 29: 554–562.
- Liu, T., Slotnick, S.D. and Yantis, S. (2004) Human MT+ mediates perceptual filling-in during apparent motion. *Neuroimage*, 21(4): 1772–1780.
- Livingstone, M.S. and Hubel, D.H. (1987) Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J. Neurosci.*, 7: 3416–3468.
- Lorenceau, J. and Shiffrar, M. (1992) The role of terminators in motion integration across contours. *Vision Res.*, 32: 263–273.
- Lu, Z.L. and Sperling, G. (1995) The functional architecture of human visual motion perception. *Vision Res.*, 35(19): 2697–2722.
- Lu, Z.L. and Sperling, G. (2001) Three-systems theory of human visual motion perception: review and update. *J. Opt. Soc. Am. A Opt. Image. Sci. Vis.*, 18(9): 2331–2370.
- Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R. and Tootell, R.B. (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci.*, 92(18): 8135–8139.
- Malach, R., Grill-Spector, K., Kushnir, T., Edelman, S. and Itzhak, Y. (1998) Rapid shape adaptation reveals position and size invariance. *Neuroimage*, 7: S43.
- Mareschal, I. and Baker, C.L. (1998) Temporal and spatial response to second-order stimuli in cat A18. *J. Neurophysiol.*, 80: 2811–2823.
- Mareschal, I. and Baker, C.L. (1999) Cortical processing of second-order motion. *Vis. Neurosci.*, 16: 527–540.
- Marr, D. (1982) *Vision*. W. H. Freeman and Co., New York.
- Mather, G. and West, S. (1993) Evidence for second-order motion detectors. *Vision Res.*, 33(8): 1109–1112.
- Mendola, J.D., Dale, A.M., Fischl, B., Liu, A.K. and Tootell, R.B.H. (1999) The representation of real and illusory contours in human cortical visual areas revealed by fMRI. *J. Neurosci.*, 19: 8560–8572.
- Mikami, A., Newsome, W.T. and Wurtz, R.H. (1986) Motion selectivity in macaque visual cortex: II. Spatiotemporal range of directional interactions in MT and V1. *J. Neurophysiol.*, 55: 1328–1339.

- Moore, C. and Engel, S.A. (2001) Neural response to perception of volume in the lateral occipital complex. *Neuron*, 29: 277–286.
- Moutoussis, K., Keliris, G., Kourtzi, Z. and Logothetis, N. (2005) A binocular rivalry study of motion perception in the human brain. *Vision Res.*, 45(17): 2231–2243.
- Muckli, L., Kriegeskorte, N., Lanfermann, H., Zanella, F.E., Singer, W. and Goebel, R. (2002) Apparent motion: event-related functional magnetic resonance imaging of perceptual switches and States. *J. Neurosci.*, 22: RC219.
- Murray, S.O., Olshausen, B.A. and Woods, D.L. (2003) Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cereb. Cortex*, 13: 508–516.
- Nakayama, K. and Silverman, G.H. (1988) The aperture problem I. Perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Res.*, 28(6): 739–746.
- Navon, D. (1976) Irrelevance of figural identity for resolving ambiguities in apparent motion. *J. Exp. Psychol. Hum. Percept. Perform.*, 2: 130–138.
- Nowlan, S.J. and Sejnowski, T.J. (1995) A selection model for motion processing in area MT of primates. *J. Neurosci.*, 15(2): 1195–1214.
- O’Keefe, L.P. and Movshon, J.A. (1998) Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey. *Vis. Neurosci.*, 15(2): 305–317.
- Olavarria, J.F., DeYoe, E.A., Knierim, J.J., Fox, J.M. and Van Essen, D.C. (1992) Neural responses to visual texture patterns in middle temporal area of the macaque monkey. *J. Neurophysiol.*, 68(1): 164–181.
- Pack, C.C., Gartland, A.J. and Born, R.T. (2004) Integration of Contour and terminator signals in visual area MT of alert macaque. *J. Neurosci.*, 24(13): 3268–3280.
- Pack, C.C. and Born, R.T. (2001) Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409(6823): 1040–1042.
- Pack, C.C., Livingstone, M.S., Duffy, K.R. and Born, R.T. (2003) End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. *Neuron*, 39(4): 671–680.
- Petersik, J.T. (1995) A comparison of varieties of “second-order” motion. *Vision Res.*, 35(4): 507–517.
- Ramachandran, V.S., Ginsburg, A.P. and Anstis, S.M. (1983) Low spatial frequencies dominate apparent motion. *Perception*, 12: 457–461.
- Ramachandran, V.S. and Gregory, R.L. (1978) Does colour provide an input to human motion perception? *Nature*, 275: 55–56.
- Reichardt, W. (1957) Autokorrelationsauswertung als funktionsprinzip des zentralnervensystems. *Z. Naturforsch.*, 12b: 447–457.
- Reichardt, W. (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In: Rosenblith, W.A. (Ed.), *Sensory Communication*. Wiley, New York.
- Rensink, R.A., O’Regan, J.K. and Clark, J.J. (1997) To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.*, 8: 368–373.
- Rensink, R.A. (2000) Visual search for change: a probe into the nature of attentional processing. *Vis. Cogn.*, 7: 345–376.
- Schiller, P. (1991) Parallel pathways in the visual system: their role in perception at isoluminance. *Neuropsychologia*, 29(6): 433–441.
- Schira, M.M., Fahle, M., Donner, T.H., Kraft, A. and Brandt, S.A. (2004) Differential contribution of early visual areas to the perceptual process of contour processing. *J. Neurophysiol.*, 91(4): 1716–1721.
- Seiffert, A.E. and Cavanagh, P. (1998) Position-based motion perception for color and texture stimuli: Effects of contrast and speed. *Vision Res.*, 39(25): 4172–4185.
- Seiffert, A.E., Somers, D.C., Dale, A.M. and Tootell, R.B. (2003) Functional MRI studies of human visual motion perception: texture, luminance, attention and after-effects. *Cereb. Cortex*, 13(4): 340–349.
- Shiffrar, M., Li, X. and Lorenceau, J. (1995) Motion integration across differing image features. *Vision Res.*, 35(15): 2137–2146.
- Shimojo, S., Silverman, G.H. and Nakayama, K. (1989) Occlusion and the solution to the aperture problem for motion. *Vision Res.*, 29: 619–626.
- Smith, A.T., Greenlee, M.W., Singh, K.D., Kraemer, F.M. and Hennig, J. (1998) The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging, fMRI. *J. Neurosci.*, 18(10): 3816–3830.
- Solomon, J.A. and Sperling, G. (1994) Full-wave and half-wave rectification in second-order motion perception. *Vision Res.*, 34(17): 2239–2257.
- Stelmach, L.B. and Herdman, C.M. (1991) Directed attention and perception of temporal order. *J. Exp. Psychol. Hum. Percept. Perform.*, 17(2): 539–550.
- Sternberg, S. and Knoll, R.L. (1973) The perception of temporal order: fundamental issues and a general model. In: Kornblum, S. (Ed.) *Attention and Performance*, Vol. IV. Academic Press, New York, pp. 629–685.
- Stone, J.V. (1999) Object recognition: view-specificity and motion-specificity. *Vision Res.*, 39: 4032–4044.
- Stoner, G.R. and Albright, T.D. (1992) Motion coherency rules are form-cue invariant. *Vision Res.*, 32(3): 465–475.
- Stoner, G.R. and Albright, T.D. (1996) The interpretation of visual motion: evidence for surface segmentation mechanisms. *Vision Res.*, 36(9): 1291–1310.
- Titchener, E.B. (1908) *Lecture on the Elementary Psychology of Feeling and Attention*. McMillan, New York.
- Tootell, R.B., Mendola, J.D., Hadjikhani, N.K., Ledden, P.J., Liu, A.K., Reppas, J.B., Sereno, M.I. and Dale, A.M. (1997) Functional analysis of V3A and related areas in human visual cortex. *J. Neurosci.*, 17(18): 7060–7078.
- Treisman, A.M. and Gelade, G.A. (1980) Feature-integration theory of attention. *Cognit. Psychol.*, 12(1): 97–136.
- Treisman, A. and Gormican, S. (1988) Feature analysis in early vision: evidence from search asymmetries. *Psychol. Rev.*, 95(1): 15–48.
- Troncoso, X.G., Macknik, S.L. and Martinez-Conde, S. (2005) Novel visual illusions related to Vasarely’s ‘nested squares’ show that corner salience varies with corner angle. *Perception*, 34(4): 409–420.

- Tse, P.U. (2006) Neural correlates of transformational apparent motion. *Neuroimage*, 31(2): 766–773.
- Tse, P.U. and Albert, M.K. (1998) Amodal completion in the absence of image tangent discontinuities. *Perception*, 27(4): 455–464.
- Tse, P.U. and Cavanagh, P. (1995) Line motion occurs after surface parsing. *Investig. Ophthalm. Vis. Sci.*, 36: S417.
- Tse, P.U., Cavanagh, P. and Nakayama, K. (1996) The roles of attention in shape change apparent motion. *Investig. Ophthalmol. Vis. Sci.*, 37: S213.
- Tse, P.U., Cavanagh, P. and Nakayama, K. (1998) The role of parsing in high-level motion processing. In: Watanabe, T. (Ed.), *High-Level Motion Processing: Computational, Neurobiological, and Psychophysical Perspectives*. MIT Press, Cambridge, MA, pp. 249–266.
- Tse, P.U. and Logothetis, N.K. (2002) The duration of 3-d form analysis in transformational apparent motion. *Percept. Psychophys.*, 64(2): 244–265.
- Tse, P.U., Sheinberg, D.L. and Logothetis, N.K. (2003) Attentional enhancement opposite a peripheral flash revealed by change blindness. *Psychol. Sci.*, 14(2): 1–8.
- Ullman, S. (1979) *The Interpretation of Visual Motion*. MIT Press, Cambridge, MA.
- Ungerleider, L. and Mishkin, M. (1982) Two cortical visual systems. In: Ingle, D., Goodale, M. and Mansfield, R. (Eds.), *Analysis of Visual Behavior*. MIT Press, Cambridge, MA, pp. 549–586.
- Vaina, L.M., Gryzacz, N.M., Saiviroonporn, P., LeMay, M., Bienfang, D.C. and Conway, A. (2003) Can spatial and temporal motion integration compensate for deficits in local motion mechanisms? *Neuropsychologia*, 41: 1817–1836.
- Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sinaert, S., Todd, J.T. and Orban, G.A. (2002) Extracting 3D from motion: differences in human and monkey intraparietal cortex. *Science*, 298: 413–415.
- van Santen, J.P.H. and Sperling, G. (1984) Temporal covariance model of human motion perception. *J. Opt. Soc. Am. A*, 1: 451–473.
- Victor, J.D. and Conte, M.M. (1990) Motion mechanisms have only limited access to form information. *Vision Res.*, 30: 289–301.
- von der Heydt, R., Peterhans, E. and Baumgartner, G. (1984) Illusory contours and cortical neuron responses. *Science*, 224(4654): 1260–1262.
- von Grünau, M.W. (1986) A motion aftereffect for long-range stroboscopic apparent motion. *Percept. Psychophys.*, 40(1): 31–38.
- von Grünau, M. and Faubert, J. (1994) Intraattribute and interattribute motion induction. *Perception*, 23(8): 913–928.
- Wallis, G. and Bühlhoff, H. (2001) Effects of temporal association on recognition memory. *Proc. Natl. Acad. Sci. USA*, 98(8): 4800–4804.
- Watson, A.B. and Ahumada Jr., A.J. (1985) Model of human visual-motion sensing. *J. Opt. Soc. Am. A*, 2: 232–342.
- Werkhoven, P., Sperling, G. and Chubb, C. (1993) The dimensionality of texture-defined motion: a single channel theory. *Vision Res.*, 33(4): 463–485.
- Wertheimer, M. (1912) Experimentelle Studien über das Sehen von Bewegung. *Z. Psychol.*, 61: 161–265.
- Wilson, H.R., Ferrera, V.P. and Yo, C. (1992) A psychophysically motivated model for two-dimensional motion perception. *Vis. Neurosci.*, 9(1): 79–97.
- Wolfe, J.M., Yee, A. and Friedman-Hill, S.R. (1992) Curvature is a basic feature for visual search tasks. *Perception*, 21(4): 465–480.
- Zanker, J.M. (1997) Is facilitation responsible for the “motion induction” effect? *Vision Res.*, 37(14): 1953–1959.
- Zanker, J.M. and Hueppgens, I.S. (1994) Interaction between primary and secondary mechanisms in human motion perception. *Vision Res.*, 34(10): 1255–1266.
- Zhou, Y.-X. and Baker, C.L. (1993) A processing stream in mammalian visual cortex neurons for non-Fourier responses. *Science*, 261: 98–101.
- Zhou, Y.-X. and Baker, C.L. (1994) Envelope-responsive neurons in Area 17 and 18 of cat. *J. Neurophysiol.*, 72: 2134–2150.
- Zhou, Y.-X. and Baker, C.L. (1996) Spatial properties of envelope responses in Area 17 and 18 of the cat. *J. Neurophysiol.*, 75: 1038–1050.
- Zhuo, Y., Zhou, T.G., Rao, H.Y., Wang, J.J., Meng, M., Chen, M., Zhou, C. and Chen, L. (2003) Contributions of the visual ventral pathway to long-range apparent motion. *Science*, 299: 417–420.
- Zucker, S.W., Dobbins, A. and Iverson, L. (1989) Two stages of curve detection suggest two styles of visual computation. *Neural Comput.*, 1: 68–81.