The utility of shape attributes in deciphering movements of non-rigid objects

Elias H. Cohen

Graduate Center for Vision Research,
State University of New York, College of Optometry,
New York, NY, USA, &
Psychology Department, Vanderbilt University,
Nashville, TN, USA

Anshul Jain

Graduate Center for Vision Research,
State University of New York, College of Optometry,
New York, NY, USA

Qasim Zaidi

Graduate Center for Vision Research,
State University of New York, College of Optometry,
New York, NY, USA

Most moving objects in the world are non-rigid, changing shape as they move. To disentangle shape changes from movements, computational models either fit shapes to combinations of basis shapes or motion trajectories to combinations of oscillations but are biologically unfeasible in their input requirements. Recent neural models parse shapes into stored examples, which are unlikely to exist for general shapes. We propose that extracting shape attributes, e.g., symmetry, facilitates veridical perception of non-rigid motion. In a new method, identical dots were moved in and out along invisible spokes, to simulate the rotation of dynamically and randomly distorting shapes. Discrimination of rotation direction measured as a function of non-rigidity was 90% as efficient as the optimal Bayesian rotation decoder and ruled out models based on combining the strongest local motions. Remarkably, for non-rigid symmetric shapes, observers outperformed the Bayesian model when perceived rotation could correspond only to rotation of global symmetry, i.e., when tracking of shape contours or local features was uninformative. That extracted symmetry can drive perceived motion suggests that shape attributes may provide links across the dorsal–ventral separation between motion and shape processing. Consequently, the perception of non-rigid object motion could be based on representations that highlight global shape attributes.

Keywords: motion from shape, non-rigid shapes, shape representation, non-rigid motion

Citation: Cohen, E. H., Jain, A., & Zaidi, Q. (2010). The utility of shape attributes in deciphering movements of non-rigid objects. *Journal of Vision, 10*(11):29, 1–15, http://www.journalofvision.org/content/10/11/29, doi:10.1167/10.11.29.

Introduction

Tumbling, rolling, swaying, stretching, leaping, spinning, flapping, dancing, kicking, bucking, jerking, sliding, gliding, tripping, shaking, wobbling, and twirling are just some of the many motions that human observers perceive and classify effortlessly, while maintaining object identity despite shape changes. All visual motions are first parsed in the striate cortex by direction-selective cells that signal local translations (Hubel & Wiesel, 1968), making it a challenge to discover how the brain disentangles different classes of complex motion from shape deformations. However, since the motion of any complex non-rigid object consists of shape transformations in systematic rather than arbitrary sequences, several authors (e.g., Jenkins & Mataric, 2004; Troje, 2002; Yacoob & Black, 1999) have shown that an object's shapes can be encoded in a low-dimensional space. Exploiting low-dimensional representations, computer vision models have extended Tomasi and Kanade's (1992) seminal factorization solution for 3-D shape from motion to the extraction of non-rigid shapes, by using either sets of basis shapes (Torresani, Hertzmann, & Bregler, 2008) or sets of basis trajectories (Akhter, Sheikh, Khan, & Kanade, 2008). The factorization algorithms, however, require the locations of each point in each image as their input, a sequence of operations that is biologically implausible. Neural models for the perception of non-rigidly articulated human motion (Giese & Poggio, 2003) suggest that view-tuned neurons in the ventral stream provide snapshots for shapes, dorsal stream neurons match patterns for trajectories, and later motion-pattern neurons combine the two streams. Indeed, some evidence suggests that neurons in the temporal cortex encode articulated humanoid actions (Singer & Sheinberg, 2010; Vangeneugden, Pollick, & Vogels, 2009). It is, however, unlikely that the brain has stored snapshots for most deforming objects. As an alternative

model, we exploit the fact that objects generally have invariant global properties, such as symmetry, and propose that abstracted shape properties can provide the information needed to separate shape deformations from global motions. We base this proposal on the results of a new experimental method designed to study shape-motion separations for arbitrarily deforming objects undergoing rotations.

We used this method to tackle three fundamental issues in object motion perception. First, we examined how disparate local motions are combined into a coherent global percept. Since motion-sensing cells in striate cortex are generally cosine-tuned (Hawken, Parker, & Lund, 1988), the motion of each local segment of an object activates neurons with preferred directions ranging over 180°, which leads to variations in local population responses across an object's boundary. In some cases such as translating plaids and the barber-pole illusion, where global motion is perceived as a single vector, combination rules such as intersection of constraints (Movshon, Adelson, Gizzi, & Newsome, 1985) or slowest/shortest motion (Weiss, Simoncelli, & Adelson, 2002) can explain the percepts. Rigid rotations require combination rules that are not as simple as for translations (Caplovitz & Tse, 2007a; Morrone, Burr, & Vaina, 1995; Weiss & Adelson, 2000) but may conform to regularization principles, such as minimal mapping (Hildreth, 1984; Ullman, 1979), smoothest motion (Hildreth, 1984), or motion coherence (Yuille & Grzywacz, 1988). By adding different forms of dynamic shape distortions to rotation, we were able to tease apart the role that global representations play in combining local motion estimates into unitary percepts.

Second, we verified that shape representations can drive veridical motion perception without directional clues, thus going beyond the divergence of motion computation and shape analysis into dorsal and ventral neural streams, respectively (Ungerleider & Mishkin, 1982). While shape-driven motion has been demonstrated in the absence of motion-energy signals for faces rotated from one side to another (Ramachandran, Armel, Foster, & Stoddard, 1998) and changes in geometrical shapes, e.g., from squares to rectangles (Tse & Logothetis, 2002), in both of these cases, the direction of motion could be inferred from the end shapes. We used randomly generated dotted shapes with indeterminate orientations to prevent such influences on rotation perception. We also quantified human efficiency for shape-driven motion perception by comparing accuracy to the simulated performance of an optimal shape-based Bayesian decoder.

Third, we tested whether properties abstracted from complex shapes can determine perceived motion. Past work has shown that shape features, such as contour completion, relatability, convexity, and closure, can determine motion grouping through surface segmentation (McDermott & Adelson, 2004; McDermott, Weiss, & Adelson, 2001) or binding (Lorenceau & Alais, 2001), and

neuroimaging has suggested that dorsal area V3A may extract features that are tracked in motion perception (Caplovitz & Tse, 2007b). However, Fang, Kersten, and Murray (2008) and Murray, Kersten, Olshausen, Schrater, and Woods (2002) presented evidence that activation of the Lateral Occipital Complex (LOC) reduced activity in striate cortex during the percept of a grouped moving stimulus but not during a non-rigid percept of the same stimulus as independently moving features. This raises the possibility that feature extraction and global shape representations may play quite different roles in object motion perception. To isolate the role of shape representations, we tested cases where tracking of shape contours or local features was uninformative, and perceived rotation could correspond only to rotation of a global axis of symmetry.

Disentangling motion direction from shape deformations

Experiment 1: Rotation detection of non-rigid shapes

We sought to determine how specific local motions are chosen from the set of possible correspondences and integrated into global percepts. In the new method, identical dots were synchronously moved in or out along invisible spokes radiating from a center, in a manner consistent with the rotation of a single stimulus (Figure 1A, Movies 1–4). The stimuli consisted of jagged shapes formed by taking 20 identical dots evenly spaced around a circle (angular gap $\theta = 18°$) and independently varying each dot's radial distance from the central fixation along its invisible spoke (Figure 1B). For each trial, the variations were drawn from a Gaussian random distribution with zero mean and standard deviation α, the *shape amplitude*. Further, on each frame in the trial, positional noise was applied to each dot's radial component, independently sampled from a random Gaussian distribution with zero mean and a pre-set standard deviation δ, the *dynamic jitter* magnitude for that trial (Figure 1B).

In Experiment 1, shape rotation between successive frames was set equal to the angle θ, so the rotation of a circle centered at fixation would be invisible in this experiment. Given the ambiguous nature of the rotation, a dot belonging to a random shape on frame i could have been perceived as moving to either of the adjacent spokes on frame $i + 1$ or along the same spoke. As illustrated in Figure 1B, the distance between a dot at frame i was generally shorter to the dot on the same spoke at frame $i + 1$ than it was to dots on adjacent spokes. A "nearest neighbor" rule is generally accepted as dominating the perceived path of apparent motion in cases where multiple locations compete for motion correspondence (e.g.,
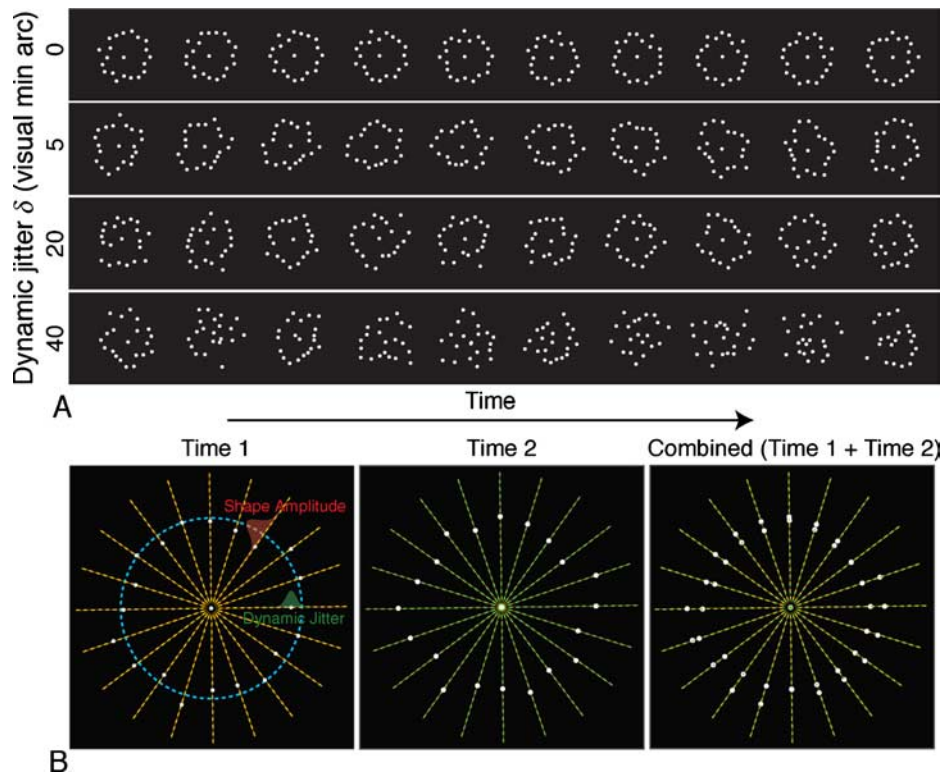
Figure 1. (A) Four representative trials from Experiment 1 at different dynamic jitter values. (B) The generation technique for Experiment 1 stimuli. Dot positions were initially set by the intersection of a circle and evenly spaced spokes. Dot variation from the circle was determined by the shape amplitude, α, for the trial and the dynamic jitter, δ, added to each frame. The first two panels depict sequential frames of the trial. The third panel superimposes the two frames to demonstrate that "nearest neighbor" motions fall along the generative spokes rather than along the veridical rotation. Dashed lines illustrate generative technique and were not visible during the experiment. Individual dots appear proportionately larger than the experimental stimuli.

Ullman, 1979). In addition, the shortest spatial excursion between two frames is also the slowest motion, which has been suggested as a governing principle in motion perception (Weiss et al., 2002). To test whether local correspondence or coherent global rotation dominates motion perception in different configurations, we measured observers' accuracy in determining the direction of rotation as a function of the standard deviation δ of *dynamic jitter*. Since, a moving form is sampled through apertures, this paradigm may seem similar to multi-slit viewing (Anstis, 2005; Kandil & Lappe, 2007; Nishida, 2004), but it is different both in intent and design. Instead of using familiar shapes to study form recognition, we used unfamiliar shapes that deform during rotation to create competition between different rules of motion combination.

Methods

There were 10 sequential frames in each trial. The 20 white dots subtended 6.6 min arc each, varied around a circular radius of 132 min arc, and were presented against a black background. *Shape amplitudes* of $\alpha = 4$ or 13 min arc were used to assign each dot a fixed radius for the

entire trial. *Dynamic jitter* was calculated independently for each dot per frame with δ set at 0, 2.5, 5, 10, 20, or 40 min arc for the trial. To vary the difficulty of the task, we used *presentation rates* of 3.5, 5.5, or 12.5 frames per second. Rotation speed was proportional to presentation rate, since all trials consisted of the same number of frames.

Stimuli were generated with a Cambridge Research Systems ViSaGe controlled by a Dell GX620 and displayed on a Sony CRT monitor with 1024×768 pixels at 120-Hz refresh rate. The observer viewed the monitor at a distance of 100 cm in a dark room with head positioned on a chin rest. For each trial, the observer used a key press to choose between clockwise and counterclockwise global rotations. No feedback was provided. Trials were presented in blocks of 36, containing each of the 36 conditions in random order. Observers viewed 20 blocks spread out over 2 days, allowing for 20 repetitions of each condition.

Data were collected from six observers, including author EC and 5 naive observers who were paid for their participation. All had normal or corrected-to-normal vision and were given prior training on the experimental task. The experiments were conducted in compliance with
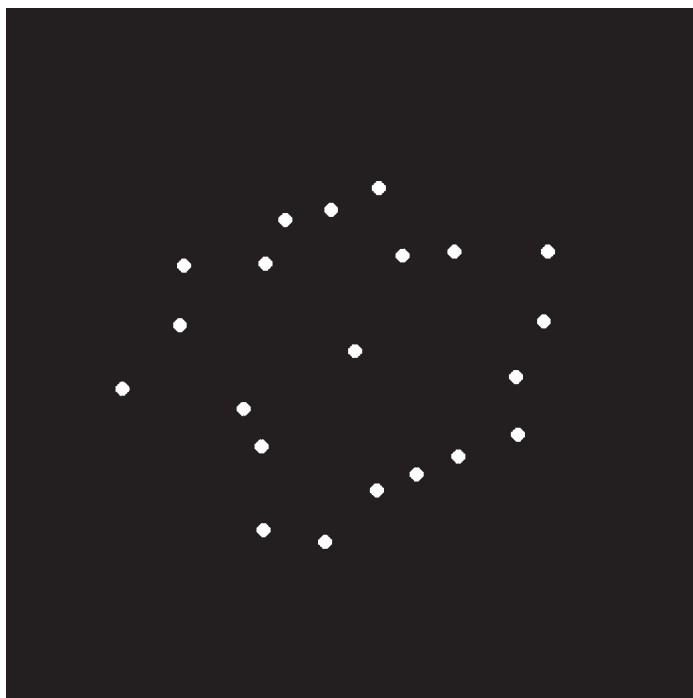
Movie 1. Representative trial from Experiment 1 at 3.5 Hz, α = 10, and dynamic jitter value (δ) = 0. Recommended viewing distance 1 m.
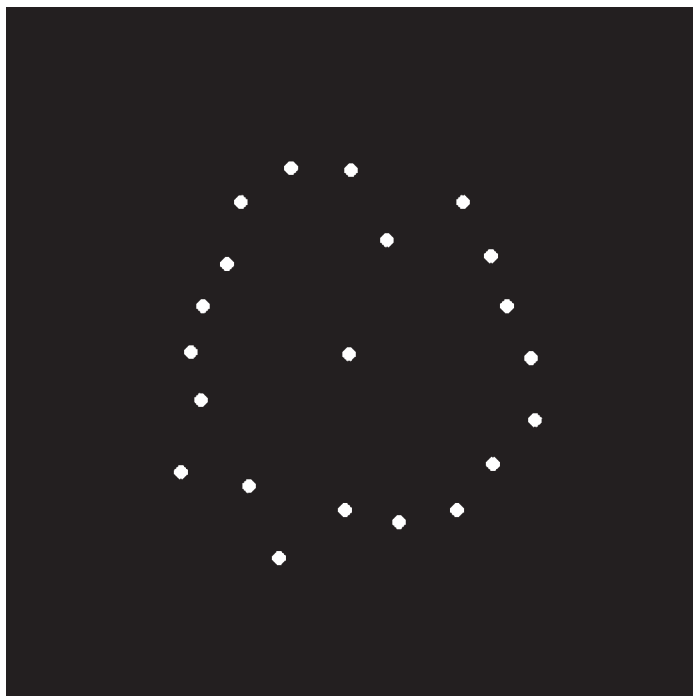
Movie 2. Representative trial from Experiment 1 at 3.5 Hz, α = 10, and dynamic jitter value (δ) = 5. Recommended viewing distance 1 m.

Movie 3. Representative trial from Experiment 1 at 3.5 Hz, α = 10, and dynamic jitter value (δ) = 20. Recommended viewing distance 1 m.

Movie 4. Representative trial from Experiment 1 at 3.5 Hz, α = 10, and dynamic jitter value (δ) = 40. Recommended viewing distance 1 m.

the standards set by the IRB at SUNY College of Optometry and observers gave their informed consent prior to participating in the experiments.

Results and modeling

Figures 2A–2C display accuracy, averaged over 6 observers, plotted against δ of the dynamic jitter, for the three presentation rates, respectively. Accuracy decreased monotonically with increasing jitter. High accuracy rates at low jitter show that global rotation was detected easily, despite competing with shorter inward and outward local motions. This shows that the global percept is not formed by combining the most salient local motions. Shapes with larger shape amplitudes (α) were significantly more resistant to dynamic jitter ($F(1, 5) = 465.3$, $p < 0.0001$) but not by a constant factor as reflected in the significant interaction between shape amplitude and jitter ($F(5, 5) = 9.9$, $p < 0.001$). An accuracy of 75% can be used as the estimated threshold for radial jitter for each shape amplitude. Thresholds generally corresponded to values

of dynamic jitter slightly greater than the shape amplitude, i.e., when the deformation of the trial shape from frame to frame was of the same order as the variations that distinguish the trial shape from a circle. This suggests that until the global rotation becomes incoherent, its percept dominates the shorter/slower local motions that indicate local expansions or contractions but do not form a coherent percept. The difference between the two presentation times was not significant, but a small improvement for less jagged shapes at faster presentation times led to a significant interaction between presentation time and shape amplitude ($F(2, 5) = 4.7$, $p < 0.05$). Johansson (1975) wrote "The eye tends to assume spatial invariance, or invariance of form, in conjunction with motion rather than variance of form without motion". The results of this experiment provide limits to Johansson's principle.

To show quantitatively that observers' rotation perception could not be explained by combining the shortest/slowest local motions, we implemented a *Nearest Neighbor Model* based on closest spatiotemporal correspondence. Each dot on frame i was matched to the nearest dot
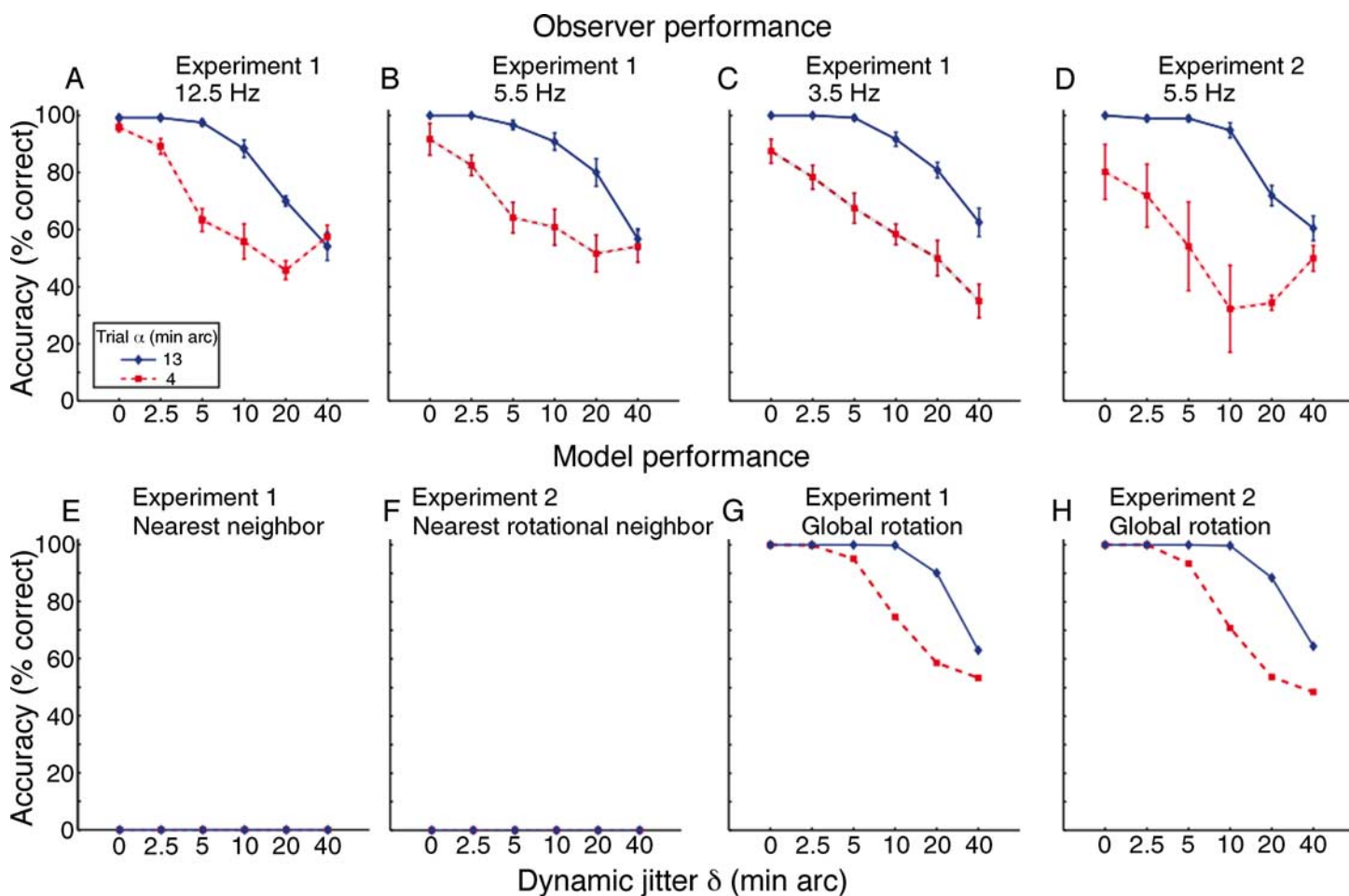

Figure 2. (A–C) Rotation discrimination in Experiment 1 averaged over 6 observers, plotted against δ of the dynamic jitter, for the three presentation rates. (D) Results from Experiment 2. (E) Predictions from *Nearest Neighbor Model* for Experiment 1. (F) Predictions from *Nearest Rotational Neighbor Model* for Experiment 2. (G) Predictions from *Global Rotation Model* for Experiment 1. (H) Predictions from *Global Rotation Model* for Experiment 2.

on frame $i + 1$. On each trial, a tally was kept of the number of clockwise, counterclockwise, and same spoke matches. The trial was classified as clockwise or counterclockwise if the majority of matches were in that direction. For the same stimuli as used in Experiment 1, the predicted percent of correct classifications is plotted against dynamic jitter in Figure 2E, showing that this model could not detect the correct direction of rotation because its input was dominated by same spoke motions.

There remain three plausible explanations for the experimental results. First, since observers are instructed to report only the direction of rotation, it is possible that they were able to ignore the radial excursions by attending only to motions from one spoke to another, and the global percept is created from the shorter/slower of the local rotary motions. Second, activation of a neural template for rotary motion, e.g., MST neurons selective for direction of rotation (Duffy & Wurtz, 1991) supersedes other motion percepts. Third, the observer matches shapes across consecutive frames and infers rotation direction from the best match. We test these possibilities in subsequent experiments and models.

Experiment 2: Strongest local motions vs. global rotation

The attention-based explanation would be consistent with the finding by Chen et al. (2008) that task-dependent spatial attention modulates neuronal firing rate in striate cortex and that response enhancement and suppression are mediated by distinct populations of neurons that differ in direction selectivity. We tested this explanation by using stimuli missing the same-spoke excursions but where the shortest/slowest local motions were in the opposite direction from the shape's rotation. In Experiment 2, each rotation was equal to 80% of the angle θ between dots (Figure 3). The shortest/slowest local motions were all individually consistent with rotation but in the direction opposite to the globally consistent rotation.

Experimental methods were identical to Experiment 1, except that the magnitude of each rotation was 80% of the distance between spokes, i.e., 14.4°. Only one presentation speed (5.5 Hz) was used. Data were collected for author EC and 3 experienced observers.

The average results are plotted in Figure 2D to allow comparison with the 5.5-Hz results from Experiment 1. The manipulation of rotation percentage made little difference to observers' accuracy in reporting rotation direction. Informal reports from observers revealed they were generally unaware of the shorter local correspondence created by partial rotation. These results argue against a combination of shortest/slowest local motions as a basis for rotation perception. If attention is involved, it may be captured by the dots moving coherently in one direction (Driver & Baylis, 1989).
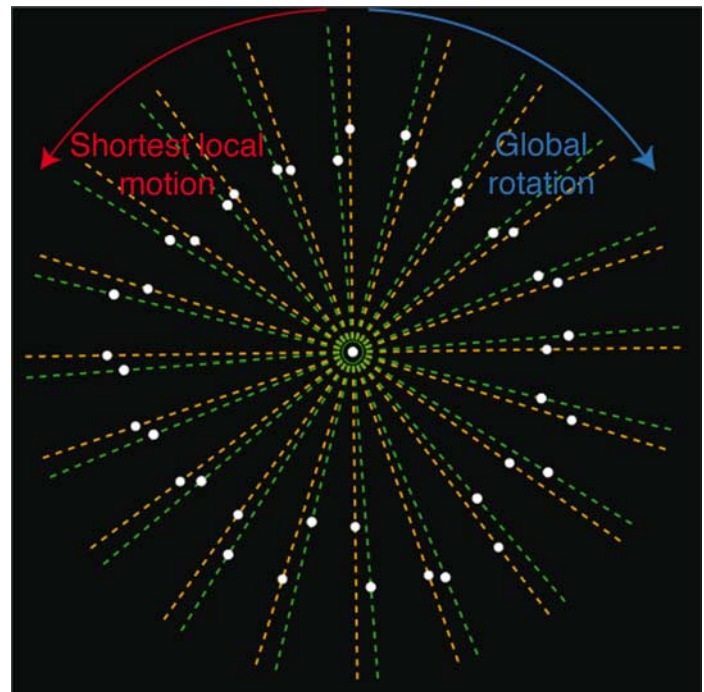

Figure 3. The correspondence conflict between successive stimulus frames in Experiment 2. Orange lines signify the shape on frame *i*. Green lines signify the shape on frame *i* + 1. For most dots, the shortest local path is in the direction opposite to the global rotation.

To provide quantitative support for this assertion, we implemented a *Nearest Rotational Neighbor Model*, which was identical to the first model, except that radial (same spoke) motions were ignored and dots were matched according to the shortest/slowest rotary motions. As would be expected, this model did better for the large amplitude shapes in Experiment 1. However, it did not predict observers' accuracy for the low amplitude shapes and failed completely on the critical test provided by Experiment 2 (Figure 2F).

Optimal global rotation model for shape-driven rotation

Two classes of neural processes could register global rotation, processes that differentiate between forms of movement, and processes that differentiate between movements of forms. The first class could be MST-like rotation templates that can signal the correct direction even if the center of rotation does not coincide with the center of the receptive field (Zhang, Sereno, & Sereno, 1993). The second class could consist of Procrustes-like processes that match shapes by discounting rotation, translation, and scaling (Mardia & Dryden, 1989). In both cases, the modeling issues are similar: how is the error estimated across each pair of frames, what function of this computed

error is used to decide the direction, how are errors accumulated across frames in a trial, and how is the direction decision made for each trial. Probability theory (Jaynes, 2003) provides optimal rules for all of these issues. In fact, at the computational level, these rules allow us to design the same optimal *Global Rotation Model* for the two distinct neural processes.

A rotating rigid shape provides a perfect fit to a rotation template and also a perfect shape match after rotation. Any error in the shape match due to dynamic jitter will be proportional to the motion deviation from a perfect rotation template; hence, we can use the same error metric for both processes. A number of error metrics have been devised for shape mismatches, but for the stimuli used in this study, a sufficient metric is to sum the squared distances between corresponding dots, after allowing for rotation, translation, and scaling. If G_i is the transition from frame i to frame $i + 1$, d_{cw}^i is the sum of squared errors for transition G_i after accounting for a clockwise rotation, and d_{cc}^i is the sum of squared errors for transition G_i after accounting for a counterclockwise rotation.

The optimal method to evaluate the plausibility ratio of the two alternative rotation directions given a particular transition is by using Bayes' theorem to relate the probability of the direction given the transition, to the likelihood that the transition occurred as a result of some rotation angle in that direction, and the prior probability of that direction (MacKay, 2003):

$$\frac{P(\text{cw}|G_i)}{P(\text{cc}|G_i)} = \frac{\sum_{\forall\theta_{cw}} P_i(\text{cw}) \cdot P(G_i|\theta_{cw})}{\sum_{\forall\theta_{cc}} P_i(\text{cc}) \cdot P(G_i|\theta_{cc})}, \qquad (1)$$

where $P_i(\text{cw})$ and $P_i(\text{cc})$ are the prior probabilities for clockwise and counterclockwise rotations (based on the experimental design, priors were set equal to 0.5). The likelihoods for stimulus transition G_i, $P(G_i|\theta_{cw})$ and $P(G_i|\theta_{cc})$, were calculated for each rotation angle independently using

$$P(G_i|\theta_k) = \frac{e^{-d_{\theta_k}^i}}{\sum_{\forall\theta} e^{-d_\theta^i}}, \qquad (2)$$

where, $d_{\theta_k}^i$ is the sum of squared errors for transition G_i after accounting for a rotation by an angle θ_k.

The rotation was considered clockwise for $0 < \theta_k < \pi$ and counterclockwise for $-\pi < \theta_k < 0$. Assuming that judgments on each transition were independent of other transitions, the plausibility ratio for each trial was taken as the product of the ratios calculated for all transitions in that trial. The outcome of the trial was taken as clockwise if the trial ratio was larger than 1.0 and as counterclockwise otherwise.

Finally, the total numbers of correct rotation decisions were tallied to get an accuracy proportion over all trails belonging to each condition. These estimates are plotted for the stimuli of Experiments 1 and 2 in Figures 2G and 2F, respectively. The *Global Rotation Model* does as well as the human observers in both experiments, suggesting that the visual system could either use a rotation template or match shapes across rotations to accomplish the task. Note that the optimal model also performs with greater accuracy for the larger shape amplitude, reflecting the easier distinctions between cc and cw shape matches as shapes depart more from the generating circle.

Human vs. model efficiency

Psychophysical results can sometimes be explained quantitatively in terms of neural properties (e.g., Cohen & Zaidi, 2007a). However, there is little information about neurons that code complex motions (Duffy & Wurtz, 1991, 1995; Oram & Perrett, 1996) or complex forms (Gallant, Braun, & Van Essen, 1993; Pasupathy & Connor, 1999; Tanaka, 1996; Tanaka, Saito, Fukada, & Moriya, 1991). Therefore, instead of a neural model that predicts rotation direction, the *Global Rotation Model* provides optimal decoding of rotations at a computational level (Marr, 1982).

To confront the model with the same problems as the human observers, calculating the shape error required the model to estimate the center and angle of rotation from the frame data. In particular, the centroid of each frame was taken as the center of rotation, creating some variability with respect to the true center. We examined whether performance would improve if we provided the shape-matching algorithm with the true center of rotation or if the center of rotation was computed as a running average of centroids for all the frames. The improvement in both cases was barely discernible. Similarly, using exact errors, analytically computed from the stimulus generation routine, to calculate likelihoods, led to only a slight improvement in performance. This was probably because the shape-matching routine is quite accurate, as reflected by the fact that the distribution of errors computed by the model was very similar to the distribution of errors computed from the combinations of Gaussian noise distributions in the shape generation routine. It is worth noting that the model's performance degraded if instead of the priors for clockwise and counterclockwise rotations being set at 0.5, the prior for each transition in the simulation was updated based on the outcome of the preceding transition.

Since human and model accuracies decrease monotonically with dynamic jitter (Figure 3) and the task used a two-alternate forced choice, we summarized the performance curves by an accuracy threshold equal to the amount of noise at the 75% accuracy level. Unlike usual thresholds, in this case a higher accuracy threshold implies that

the observer can tolerate more noise and hence performs better. The horizontal lines in Figure 4 show accuracy thresholds for human performance at three speeds for the higher shape amplitude in Experiment 1, where the shapes were more distinct. We simulated the thresholds of the optimal *Global Rotation Model* for the stimuli of Experiment 1, but instead of considering the whole shape, we considered only 1, 2, … or 20 consecutive dots, chosen randomly for each frame transition. The labels on the top of Figure 4 convert number of dots considered to percentage of available information used, which we will use as a measure of equivalent efficiency for human observers. At the slowest speed we tested, human observers performed almost as well as the model that used 18 points, i.e., at 90% of the efficiency of the optimal decoder. This implies that the human visual system includes near optimal processes for matching deforming shapes and/or for detecting rotation in the presence of strong distracting motions. The equivalent efficiency of human observers declined at the faster presentation rates. The equivalent linear speeds at these presentation rates were 3.8 and 8.6 dva/s. Since motion energy is extracted well at these speeds (Lu & Sperling, 1995; Zaidi &
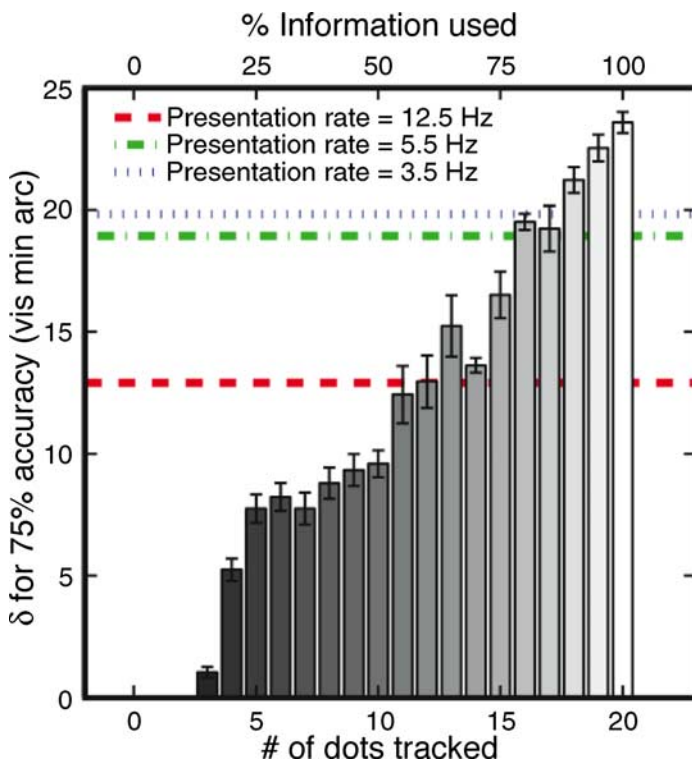

Figure 4. Seventy-five percent accuracy thresholds for human performance at three presentation rates in Experiment 1 compared with the performance of the optimal *Global Rotation Model* using subsets of dots, or equivalently percent of available information. Error bars for the model simulations are standard errors calculated from 6 independent sets of stimuli, each containing the same number of trials as the psychophysical experiments.

DeBonet, 2000), observer limitations at higher speeds may reflect the number of dots that can be used in shape or motion computations at the shorter stimulus durations.

Shape-driven motion

Experiment 3: Shape axis motion

The next challenge was to devise an experiment in which rotation direction could be determined from an abstracted shape property in conditions where shape matches and rotation templates are insufficient. We sought a salient shape attribute that could retain spatiotemporal continuity without having features that are correlated across frames. Shape axes have been considered central defining features of shape representation (Marr & Nishihara, 1978), and this may be particularly true for axes of bilateral symmetry (Wilson & Wilkinson, 2002). In the extreme case, we presented essentially unique symmetric shapes on each frame, so that dot-based correlations were absent between successive frames, but the orientations of the symmetry axes rotated consistently in one direction. Thus, only an extracted symmetry axis could provide a cue to rotation direction.

Stimuli were varied for symmetry in two ways. *Base shapes* were generated from circles to be either symmetric or asymmetric, and *dynamic jitter* was also either symmetric or asymmetric. Trials with symmetric base shapes and high levels of symmetric dynamic jitter consisted of a series of distinct symmetric shapes with a continuously turning axis (Figure 5, Movies 5–8).

In Experiment 3, all stimulus parameters from Experiment 1 were repeated for all combinations of base shape and dynamic jitter. Therefore, there were four times as many experimental conditions (2 types of base shape × 2 types of dynamic jitter). This produced a total of 144 conditions (36 × 4). Each experimental block contained all conditions in random order. Twenty blocks were run for each subject. All six participants from Experiment 1 participated in Experiment 3.

Performance accuracy in Experiment 3 as a function of dynamic jitter is plotted in Figures 6A–6C and 6E–6G for two shape amplitudes and three presentation rates. With low levels of dynamic jitter, rotations of symmetric and asymmetric shapes were detected equally. With increasing dynamic jitter, the symmetric dynamic jitter results form U-shaped functions. As dynamic jitter magnitude increased, making each sequential shape more unique from its temporal neighbors, symmetric motions remained detectable for the two slower presentation rates. Observers perceived the stimulus as rotating, despite lack of correspondence between local elements, at magnitudes of dynamic jitter that made it impossible for observers to detect asymmetric global rotation at any presentation rate. There was a significant main effect for symmetric versus
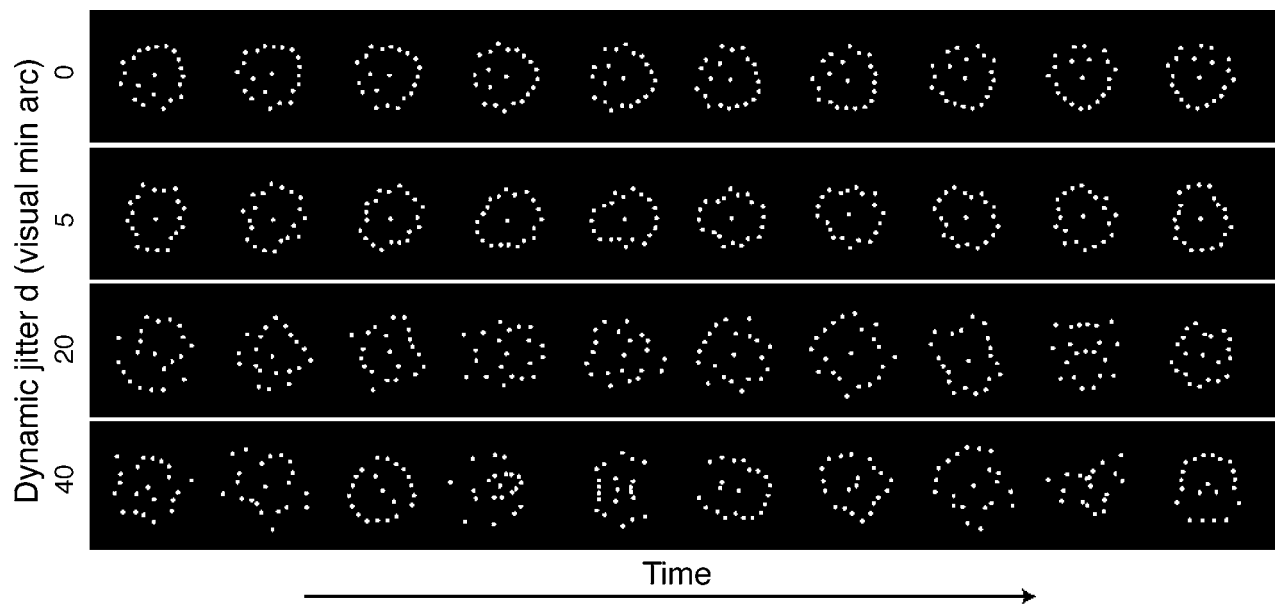
Figure 5. Four representative trials in the symmetric shape/symmetric jitter condition from Experiment 3 at four different dynamic jitter levels. Trials with high dynamic jitter appear to display a unique shape on each frame while presenting a unidirectional rotation.

asymmetric jitter ($F(1, 5) = 130.5$, $p < 0.001$) and an interaction between jitter symmetry and jitter amplitude ($F(5, 5) = 33.6$, $p < 0.001$). As shown in Figures 6D and 6H, the *Global Rotation Model* demonstrated no significant difference between symmetric and asymmetric stimuli from Experiment 3. This demonstrates that observers' performance on symmetric shapes was not due to dot-based shape or motion correlations. Comparison of human and model performances show that observers were capable of outdoing the optimal global model by extracting relevant shape attributes that are invisible to shape matching and rotation templates.

The question arises whether discerned movement of symmetric shapes is based on perceived motion or on perceived changes of orientation, a variant of the dichotomy between motion energy and feature tracking (Lu & Sperling, 1995; Zaidi & DeBonet, 2000). Phenomenologically, in these trials, unless the shapes were perceived as symmetric, the dots were seen in random motion. When symmetry was perceived, the impression was that of a shape changing during rotation but not always smoothly. Direction discernment thus required following the rotation actively, consistent with feature tracking or "conceptual motion" (Blaser & Sperling, 2008; Seiffert & Cavanagh, 1998). However, symmetry-based motion exhibits at least one signature of "perceptual motion", i.e., motion aftereffects. To test for motion aftereffects, we adapted for 30-s epochs to shapes from the Symmetric-shape + Symmetric-jitter conditions with $\alpha = 4$ and $\delta = 40$ rotated at 3.5 Hz in Experiment 3. As shown in Figure 6C, the greatest performance advantage for symmetry occurs in this condition. Two authors (AJ and QZ) reported the direction of the adapting stimulus and then judged the direction of rotation of dots forming a circle of the same average size as the adapting stimulus. The dots of the circle were rotated half the inter-spoke angle at 3.5 Hz, hence were equally likely to be seen to move in clockwise and counterclockwise directions, unless biased by motion adaptation. For the Symmetric shapes, AJ judged 30/30 of the adapting directions correctly while QZ judged 26/30 correctly. More to the point, AJ judged the motion aftereffects to be in the direction opposite to the simulated motion 26/30 times and QZ judged 22/30 times (chance performance can be rejected for both observers at $p < 0.01$). Both observers noted that unlike judging the direction of the jittered symmetric shapes, judging the direction of the aftereffects seemed effortless. These aftereffects were not an artifact, because reliable aftereffects did not result from adaptation to Asymmetric shape + Asymmetric jitter of the same size at the same speed: percent corrects for Asymmetric adapting shapes were essentially at chance, 13/30 and 15/30, as were aftereffects reported in the direction opposite to the simulated motion, 13/30 and 16/30.

## Experiment 4: Symmetry extraction durations

The results of Experiment 3 showed a decrease in accuracy for symmetric stimuli at higher presentation rates. To test whether this effect is due to the time demands of symmetry detection, Experiment 4 was designed to measure duration thresholds for perceiving the rotation of symmetric shapes. Four of the six observers (author EC and three others) were presented with rotations of either rigid asymmetric shapes or symmetric shapes with large magnitudes of symmetric dynamic jitter, at nine presentation rates ranging from 1.5 to 60.0 Hz.
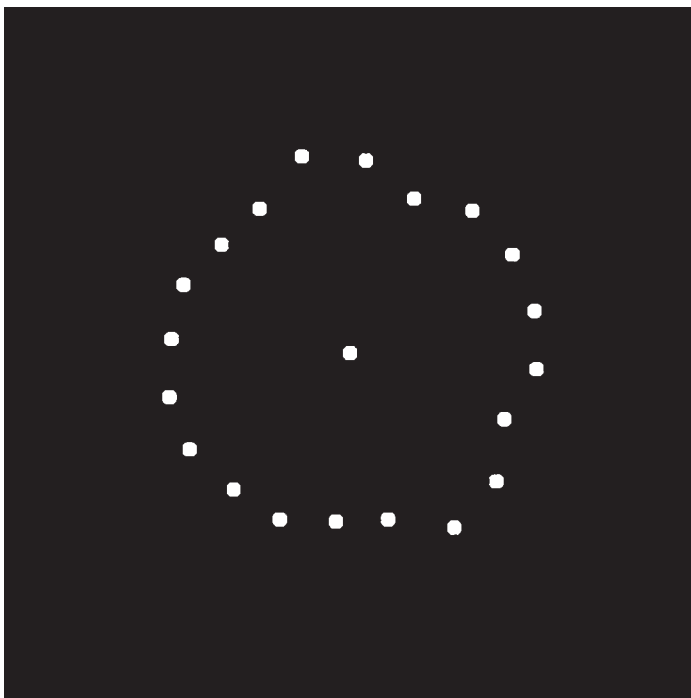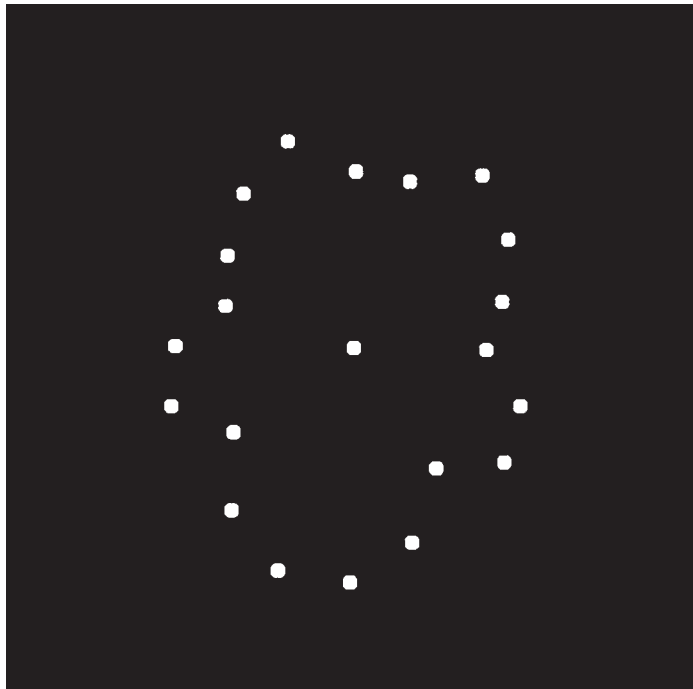
Movie 5. Representative trial from Experiment 3 at 2.5 Hz and symmetric dynamic jitter value (δ) = 0. Recommended viewing distance 1 m.
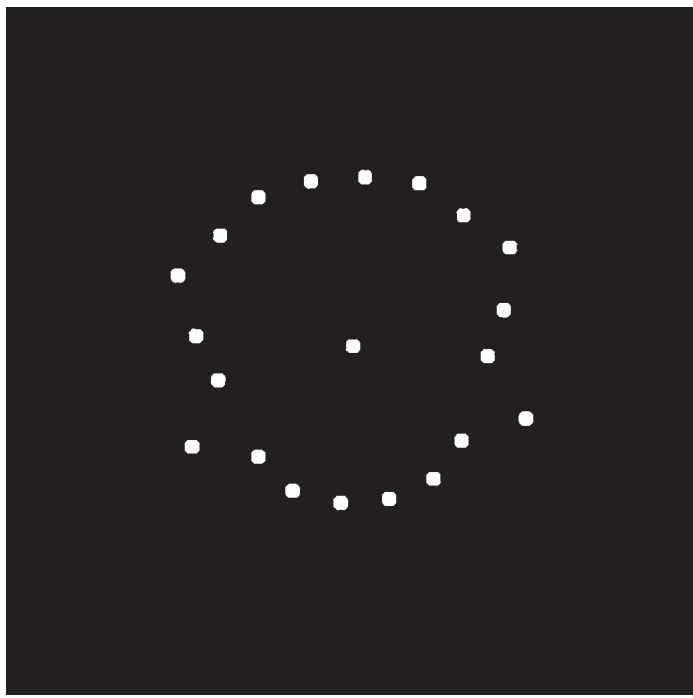
Movie 6. Representative trial from Experiment 3 at 2.5 Hz and symmetric dynamic jitter value (δ) = 5. Recommended viewing distance 1 m.
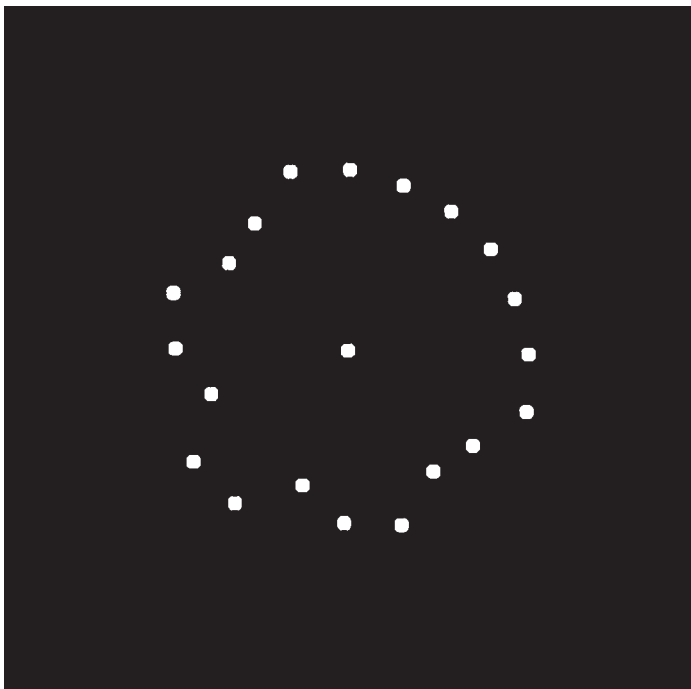
Movie 7. Representative trial from Experiment 3 at 2.5 Hz and symmetric dynamic jitter value (δ) = 20. Recommended viewing distance 1 m.

Movie 8. Representative trial from Experiment 3 at 2.5 Hz and symmetric dynamic jitter value (δ) = 40. Recommended viewing distance 1 m.
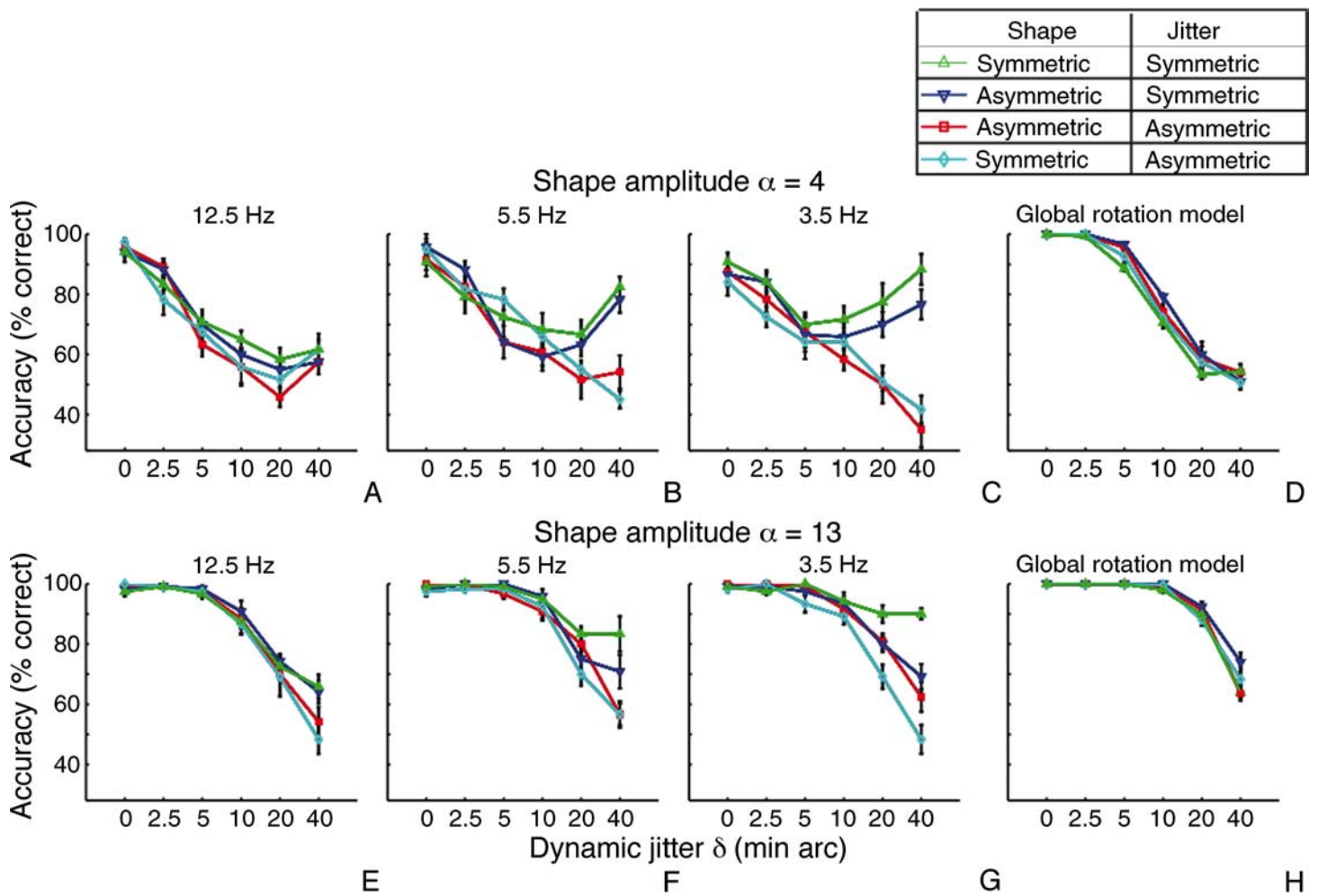
Figure 6. (A–C and E–G) Results from Experiment 3 for symmetric and asymmetric random shapes. (D, H) Accuracy estimates from the *Global Rotation Model* for the two shape amplitudes.
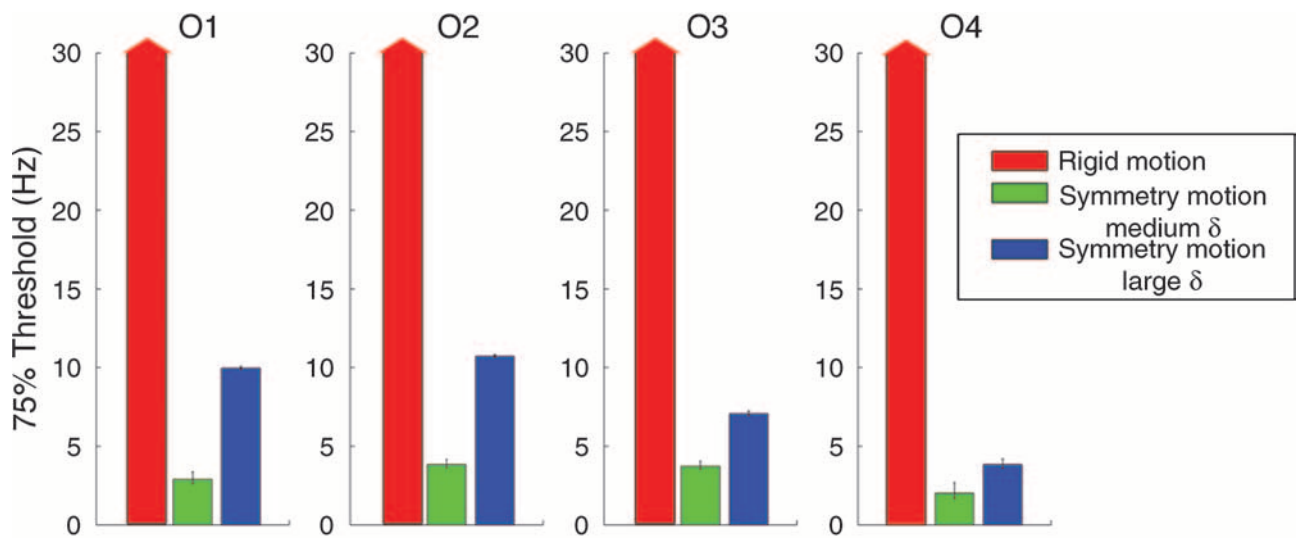

Figure 7. Results from Experiment 4 for each of four observers (O1–O4). Red arrows for rigid motion stimuli indicate that performance was above threshold at all presented rates. Green bars show results for shape amplitude $\delta = 20$ and blue bars for $\delta = 40$.

Figure 7 shows the 75% presentation rate thresholds estimated from psychometric functions for accuracy (Wichmann & Hill, 2001). Observer performance for rigid motion trials was above threshold for all presentation rates. Average threshold for detecting symmetry axis rotation was 7.9 Hz ($SE = 1.6$) for jitter amplitude $\delta = 40$ and 3.1 Hz ($SE = 0.41$) for $\delta = 20$. Symmetry perception is possible in intervals as short as 50 ms under ideal conditions (Julesz, 1971) but can take considerably longer for complex stimuli (Cohen & Zaidi, 2007b) and non-vertical orientations (Barlow & Reeves, 1979). The presentation rate threshold for the larger δ translates into a frame duration threshold of 125 ms. This result suggests that observers need presentation durations compatible with symmetry extraction to detect rotation of the symmetry axis.

Discussion

The results of this study have a number of implications for understanding object motion perception. The results of Experiments 1 and 2 show that human observers are almost as efficient at detecting the rotation of arbitrarily deforming objects as the optimal statistical decoder. Despite competition from stronger local motion signals, global rotation is discriminated easily and accurately in the presence of even high levels of positional noise. The random jitter in this method may be seen as reflecting a worst-case scenario for motion of deforming objects. In real-world deformations, local motions will be more systematic and there is likely to be a higher shape correlation across successive frames. Human observers may thus be expected to do better than a shape-matching model for most natural deformations, especially those that present extra information like elongations.

The results of Experiments 3 and 4 show that observers can use symmetry axes to infer motion direction, even when form or motion information from the contour is uninformative. Stable appearance descriptors can be a tremendous aid in the critical task of object recognition. For example, shape is the geometric descriptor that is invariant to translation, rotation, and scaling (Kendall, Barden, Carne, & Le, 1999). It has been suggested that by detecting and encoding an object's shape structure, the visual system may form a robust object representation that is stable across changes in viewing conditions and efficiently characterizes spatially distributed information (Biederman, 1987; Hoffman & Richards, 1984; Marr & Nishihara, 1978). Some retinal projections of 3-D objects, however, introduce shape distortions that the visual system is unable to discount (Griffiths & Zaidi, 1998, 2000). In addition, many objects are articulated, plastic, or elastic, so a rigid shape description is insufficient. In such cases, invariant shape properties like symmetry may aid in recreation and recognition of volumetric shapes from images (Pizlo, 2008). In fact, bilateral symmetry is widespread in natural and man-made objects (Tyler, 1996), and axes based on local symmetry have been found to be useful in representing shapes (Blum, 1973; Feldman & Singh, 2006; Leyton, 1992; Marr & Nishihara, 1978). Our results show that in identifying the path of a moving object, where disparate and spatially distributed local motion signals need to be combined (Hildreth, 1984; Ullman, 1979; Yuille & Grzywacz, 1988), the visual system may benefit from representation of object structures such as axes of symmetry. Symmetry was chosen as the property examined in these studies due to its separability from local contour properties, but it is possible that other abstract properties related to the contour may also benefit motion processing.

The motivation for this study was to identify the influence of form processing on the perception of object motion. The results of this study identify a number of such influences. First, the selection of local motion information for combination into a global percept depends not on the relative strengths of local directional signals but on the most plausible/coherent global motion (note that the dot motions are compatible with rotation in either direction if it is coupled with sufficient shape distortions). The global percept may be a result of the activation of a rotation template overriding an incoherent percept formed by the strongest local motion signals, or it may result from a process that infers the direction of motion as that which gives the strongest shape correlation across frames. The latter generalizes the basic idea behind Reichardt's (1961) detector to object motions more complex than translation (Lu & Sperling, 1995). Second, the shapes we used in the first two experiments had no inherent orientation, so the inference of rotation from one frame to the other could not have been based on any intrinsic quality of the shapes but solely on the post-rotation point-wise correlation or the most coherent global motion. Third, for these arbitrary shapes, the high efficiency of human motion perception could be due to either just motion or just shape information or both. The above threshold accuracy for motion detection of the dynamically varying symmetric shapes, however, cannot be based on any rotation-sensitive neuron fed by motion-energy signals, or on correlating contours of the shape, so it must be based on shape representations that explicitly label the extracted symmetry axes. It is possible that shape representations that include abstracted attributes may provide solutions to a broad range of problems in conditions where contours or features create ambiguous percepts. A general model for the perception of object motion should thus include representation processes that highlight shape attributes.

Acknowledgments

Commercial relationships: none.
Corresponding author: Qasim Zaidi.
Email: qz@sunyopt.edu.
Address: Graduate Center for Vision Research, State University of New York, College of Optometry, 33 W 42nd Street, New York, NY 10036, USA.

References

Akhter, I., Sheikh, Y. A., Khan, S., & Kanade, T. (2008). *Nonrigid structure from motion in trajectory space*. Paper presented at the Neural Information Processing Systems.

Anstis, S. (2005). Local and global segmentation of rotating shapes viewed through multiple slits. *Journal of Vision, 5*(3):4, 194–201, http://www.journalofvision.org/content/5/3/4, doi:10.1167/5.3.4. [PubMed] [Article]

Barlow, H. B., & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research, 19,* 783–793.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Blaser, E., & Sperling, G. (2008). When is motion "motion"? *Perception, 37,* 624–627.

Blum, H. (1973). Biological shape and visual science. I. *Journal of Theory of Biology, 38,* 205–287.

Caplovitz, G. P., & Tse, P. U. (2007a). Rotating dotted ellipses: Motion perception driven by grouped figural rather than local dot motion signals. *Vision Research, 47,* 1979–1991.

Caplovitz, G. P., & Tse, P. U. (2007b). V3A processes contour curvature as a trackable feature for the perception of rotational motion. *Cerebral Cortex, 17,* 1179–1189.

Chen, Y., Martinez-Conde, S., Macknik, S. L., Bereshpolova, Y., Swadlow, H. A., & Alonso, J. M. (2008). Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nature Neuroscience, 11,* 974–982.

Cohen, E. H., & Zaidi, Q. (2007a). Fundamental failures of shape constancy resulting from cortical anisotropy. *Journal of Neuroscience, 27,* 12540–12545.

Cohen, E. H., & Zaidi, Q. (2007b). Salience of mirror symmetry in natural patterns [Abstract]. *Journal of Vision, 7*(9):970, 970, http://www.journalofvision.org/content/7/9/970, doi:10.1167/7.9.970.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 448–456.

Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology, 65,* 1329–1345.

Duffy, C. J., & Wurtz, R. H. (1995). Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *Journal of Neuroscience, 15,* 5192–5208.

Fang, F., Kersten, D., & Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision, 8*(7):2, 1–9, http://www.journalofvision.org/content/8/7/2, doi:10.1167/8.7.2. [PubMed] [Article]

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 18014–18019.

Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science, 259,* 100–103.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4,* 179–192.

Griffiths, A. F., & Zaidi, Q. (1998). Rigid objects that appear to bend. *Perception, 27,* 799–802.

Griffiths, A. F., & Zaidi, Q. (2000). Perceptual assumptions and projective distortions in a three-dimensional shape illusion. *Perception, 29,* 171–200.

Hawken, M. J., Parker, A. J., & Lund, J. S. (1988). Laminar organization and contrast sensitivity of direction-selective cells in the striate cortex of the Old World monkey. *Journal of Neuroscience, 8,* 3541–3548.

Hildreth, E. C. (1984). *Measurement of visual motion*. Cambridge, MA: MIT Press.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition, 18,* 65–96.

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

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge, UK: Cambridge University Press.

Jenkins, O. C., & Mataric, M. J. (2004). *A spatio-temporal extension to isomap nonlinear dimension reduction*. Paper presented at the 21st International Conference on Machine Learning, Banff, Canada.

Johansson, G. (1975). Visual motion perception. *Science America, 232,* 76–88.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: The University of Chicago Press.

Kandil, F. I., & Lappe, M. (2007). Spatio-temporal interpolation is accomplished by binocular form and motion mechanisms. *PLoS One, 2,* e264.

Kendall, D. G., Barden, D., Carne, T. K., & Le, H. (1999). *Shape and shape theory*. Hoboken, NJ: John Wiley & Sons.

Leyton, M. (1992). *Symmetry, causality, mind*. Cambridge, MA: MIT Press.

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

Lu, Z. L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research, 35,* 2697–2722.

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK: Cambridge University Press.

Mardia, K. V., & Dryden, I. L. (1989). The statistical analysis of shape data. *Biometrika, 76,* 271–281.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biology Science, 200,* 269–294.

McDermott, J., & Adelson, E. H. (2004). The geometry of the occluding contour and its effect on motion interpretation. *Journal of Vision, 4*(10):9, 944–954, http://www.journalofvision.org/content/4/10/9, doi:10.1167/4.10.9. [PubMed] [Article]

McDermott, J., Weiss, Y., & Adelson, E. H. (2001). Beyond junctions: Nonlocal form constraints on motion interpretation. *Perception, 30,* 905–923.

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

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattas, & C. G. Gross (Eds.), *Pattern recognition mechanisms* (pp. 117–151). Rome, Italy: Vatican Press.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 15164–15169.

Nishida, S. (2004). Motion-based analysis of spatial patterns by the human visual system. *Current Biology, 14,* 830–839.

Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of Neurophysiology, 76,* 109–129.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology, 82,* 2490–2502.

Pizlo, Z. (2008). *3D shape: Its unique place in visual perception*. Cambridge, MA: MIT Press.

Ramachandran, V. S., Armel, C., Foster, C., & Stoddard, R. (1998). Object recognition can drive motion perception. *Nature, 395,* 852–853.

Reichardt, H. P. (1961). Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), *Principles of sensory communications* (pp. 303–317). Hoboken, NJ: John Wiley & Sons.

Seiffert, A. E., & Cavanagh, P. (1998). Position displacement, not velocity, is the cue to motion detection of second-order stimuli. *Vision Research, 38,* 3569–3582.

Singer, J. M., & Sheinberg, D. L. (2010). Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. *Journal of Neuroscience, 30,* 3133–3145.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Reviews Neuroscience, 19,* 109–139.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66,* 170–189.

Tomasi, C., & Kanade, T. (1992). Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision, 9,* 18.

Torresani, L., Hertzmann, A., & Bregler, C. (2008). Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30,* 15.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision, 2*(5):2, 371–387, http://www.journalofvision.org/content/2/5/2, doi:10.1167/2.5.2. [PubMed] [Article]

Tse, P. U., & Logothetis, N. K. (2002). The duration of 3-D form analysis in transformational apparent motion. *Perception & Psychophysics, 64,* 244–265.

Tyler, C. W. (1996). *Human symmetry perception and its computational analysis*. Utrecht, The Netherlands: VSP Press.

Ullman, B. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavioral* (pp. 549–586). Cambridge, MA: MIT Press.

Vangeneugden, J., Pollick, F., & Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. *Cerebral Cortex, 19,* 593–611.

Weiss, Y., & Adelson, E. H. (2000). Adventures with gelatinous ellipses—Constraints on models of human motion analysis. *Perception, 29,* 543–566.

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

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313.

Wilson, H. R., & Wilkinson, F. (2002). Symmetry perception: A novel approach for biological shapes. *Vision Research, 42,* 589–597.

Yacoob, Y., & Black, M. J. (1999). Parametrized modeling and recognition of activities. *Computer Vision and Image Understanding, 73,* 232–247.

Yuille, A. L., & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature, 333,* 71–74.

Zaidi, Q., & DeBonet, J. S. (2000). Motion energy versus position tracking: Spatial, temporal, and chromatic parameters. *Vision Research, 40,* 3613–3635.

Zhang, K., Sereno, M. I., & Sereno, M. E. (1993). Emergence of position-independent detectors of sense of rotation and dilation with Hebbian learning: An analysis. *Neural Computation, 5,* 597–612.