Perceptual segmentation and the perceived orientation of dot clusters: The role of robust statistics

Elias H. Cohen
Department of Psychology, Rutgers University, New Brunswick, NJ, USA, &
Department of Vision Sciences, SUNY, State College of Optometry, New York, NY, USA

Manish Singh
Department of Psychology and Center for Cognitive Science, Rutgers University, New Brunswick, NJ, USA

Laurence T. Maloney
Department of Psychology and Center for Neural Science, New York University, New York, NY, USA

We investigated perceptual segmentation in the context of a perceived-orientation task. Stimuli were dot clusters formed by the union of a large elliptical sub-cluster and a secondary circular sub-cluster. We manipulated the separation between the two sub-clusters, their common dot density, and the size of the secondary sub-cluster. As the separation between sub-clusters increased, the orientation perceived by observers shifted gradually from the global principal axis of the entire cluster to that of the main sub-cluster alone. Thus, with increasing separation, the dots within the secondary sub-cluster were assigned systematically lower weights in the principal-axis computation. In addition, this shift occurred at smaller separations for higher dot densities—consistent with the idea that reliable segmentation is possible with smaller separations when the dot density is high. We propose that the visual system employs a robust statistical estimator in this task and that data points are weighted differentially based on the likelihood that they arose from a separate generative process. However, unlike in standard robust estimation, weights based on residuals are insufficient to characterize human segmentation. Rather, these must be computed based on more comprehensive generative models of dot clusters.

Keywords: orientation perception, principal axis, perceptual grouping, probabilistic segmentation, robust statistics

Citation: Cohen, E. H., Singh, M., & Maloney, L. T. (2008). Perceptual segmentation and the perceived orientation of dot clusters: The role of robust statistics. *Journal of Vision, 8*(7):6, 1–13, http://journalofvision.org/8/7/6/, doi:10.1167/8.7.6.

Introduction

The estimation of any visual property (e.g., position, orientation, motion) requires that information be pooled over an extended image region. Such pooling is only useful, however, if the selected region is appropriate to the estimation task. In estimating the motion of an object, for instance, pooling is necessary because of the inherent ambiguity of local motion signals (the *aperture problem*). However, such pooling can also yield highly biased estimates if motion signals from two different objects are mixed—e.g., if along with the moving object of interest, the pooling also includes motion signals from a stationary background or a nearby object moving in the opposite direction (e.g., McDermott & Adelson, 2004; McDermott, Weiss, & Adelson, 2001). Similarly, object localization (e.g., for guiding saccades) requires that the visual information for one object be segregated from visual information corresponding to other objects before an estimate of location is computed (Cohen, Schnitzer, Gersch, Singh, & Kowler, 2007; Denisova, Singh, & Kowler, 2006; Melcher & Kowler, 1999; Vishwanathan, Kowler, & Feldman, 2000). Thus, the estimation of any visual property presupposes perceptual segmentation, the division into "perceptual groups."

Perceptual segmentation includes a variety of phenomena, including figure-ground segmentation (Driver & Baylis, 1996; Koffka, 1935; Peterson & Gibson, 1994; Rubin, 1915/1958), segmentation of dot arrays into perceptual groups (Kubovy & Wagemans, 1995; Rock & Brosgole, 1964; Wertheimer, 1923), segmenting objects into component parts (Biederman, 1987; De Winter & Wagemans, 2006; Hoffman & Richards, 1984; Marr & Nishihara, 1978; Singh & Hoffman, 2001), etc. Here we focus on the segmentation of dot clusters into potentially two sub-clusters. Perceptual segmentation is often treated as a binary process, i.e., a region or set of dots is either segmented or not—it is either one "perceptual unit" or two. However, we suggest it may be more fruitful to think of segmentation as graded. It is plausible that the visual system represents degrees of "belief" concerning perceptual segmentation in any given instance, depending on the strength of evidence from multiple sources. Such a proposal is consistent with recent work on contour integration and interpolation, which models the grouping

doi: 10.1167/8.7.6 Received October 18, 2007; published May 23, 2008 ISSN 1534-7362

of image elements into the representation of a single extended contour as a probabilistic process (Elder & Goldberg, 2002; Feldman, 1997, 2001; Geisler, Perry, Super, & Gallogly, 2001; Hon, Maloney, & Landy, 1997; Singh & Fulvio, 2005; Warren, Maloney, & Landy, 2002, 2004).

In the current study, we examine the influence of perceptual segmentation on the visual estimation of a global property when there is uncertainty concerning perceptual segmentation. We use the context of perceived orientation of dot clusters, comprising a large "main" sub-cluster and a smaller "secondary" sub-cluster that could either be perceived as separate or as part of the main cluster (see Figure 1a).

We examine the visual estimation of global orientation as a means of measuring probabilistic segmentation: The stronger the evidence that the secondary sub-cluster is a separate perceptual unit, the weaker should be that sub-cluster's influence on the perceptual estimate of global orientation.

Perceived orientation of dot clusters and shapes

Previous work has shown that the perceived orientation of a dot cluster is well predicted by its first principal axis. This is the line that minimizes the sum of squared distances to the dots, where the distances are measured perpendicular to the candidate line[1] (Lánský, Yakimoff, & Radil, 1987; Lánský, Yakimoff, Radil, & Mitrani, 1989; Yakimoff, 1981; Yodogawa, 1985). The visual system's reliance on the principal axis has been demonstrated for dot clusters sampled from different distributions—including a uniform distribution on rectangular regions (Yakimoff, 1981) and bivariate Gaussian distributions (Lánský et al., 1987, 1989). There are also, however, secondary influences, such as a small bias toward the cardinal and the ±45° directions (Lánský et al., 1989). Not surprisingly, the precision of observers' orientation estimates decreases with increasing spread of the dots (i.e., decrease in correlation). However, Lánský et al. (1987) found that observers' performance in estimating orientation remained above chance even for correlations as low as 10%.

Previous work on the perceived orientation of 2D shapes is also consistent with a principal-axis computation, even though the results are not always articulated in these terms. Li and Westheimer (1997) speak of the "implicit orientation" of shapes; for instance, a shape such as an X is perceived as being vertical even though all of its edges are oblique. They showed that the implicit orientation of shapes exhibits many of the same characteristics as the explicit orientation of a line segment (notably, the oblique effect; see also Liu, Dijkstra, & Oomes, 2002). They thus reasoned that the computation of global shape orientation may be based on mechanisms closely related to those involved in the computation of line orientation. For our current purposes, it is noteworthy that, in their examples (e.g., an X or an ellipse), the implicit orientation corresponds precisely to the orientation of the principal axis.

Burbeck and Zauberman (1997) studied perceived orientation on rectangular stimuli whose longer sides were modulated sinusoidally with waves of varying frequency and relative phase. Based on their results, Burbeck and Zauberman argued against a model based on averaging the edge orientations of the shape's bounding contour. They proposed instead that a shape's perceived orientation is mediated by its skeletal description, as given by the medial-axis transform[2] (Blum, 1973). Although the medial axis may indeed play a role, it is again noteworthy that Burbeck and Zauberman's results are also consistent with a principal-axis computation—a possibility they did not consider (see Cohen & Singh, 2006).

Oomes and Dijkstra (2002) investigated the perceived orientation of ellipsoids and other symmetric three-dimensional shapes. Contrary to results with 2D shapes, they found systematic deviations of observers' settings from the 3D principal axis. However, a decomposition of

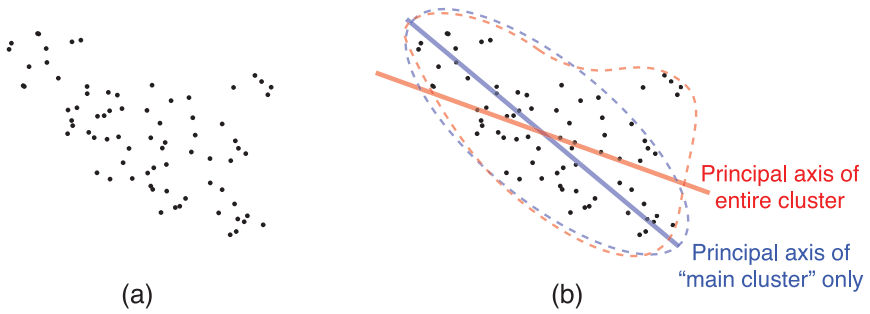

Figure 1. (a) A stimulus that can be interpreted either as a single dot cluster or as a composite consisting of a large elliptical sub-cluster and a small, separate sub-cluster. (b) Predictions of orientation based on a principal-axis computation based either on the entire cluster as a whole (shown in red), or based on the main sub-cluster alone, ignoring the smaller sub-cluster (shown in blue).

observers' settings into tilt (orientation in the image plane) and slant (orientation in depth) revealed that these deviations arose almost entirely from the slant component. Their results thus showed that while observers underestimate the orientation in depth of 3D objects (how much the object is slanted away from the frontal plane), they are nevertheless proficient at using the principal axis of the 2D projection of the object in estimating its orientation in the image plane.

Cohen and Singh (2006) studied perceived orientation of multi-part 2D shapes comprising a small part protruding out of a large "base." Their results showed systematic deviations of perceived orientation from the shapes' principal axis. The magnitude of this deviation increased with the salience of the part's boundaries (based on the curvature magnitude at the points of negative minima of curvature bounding the part; see Hoffman & Singh, 1997). As the boundary points became sharper, and the part more distinct, observers' orientation estimates approached the principal axis of the base part, suggesting that the small protruding part was being largely ignored (see Figure 2). These results suggest that part segmentation is a graded process that is reflected in the computation of global properties of a shape composed of multiple parts.

Robust statistics

Robust statistics is a collection of mathematical methods concerned with developing statistical estimators that are little affected by small failures of distributional assumptions (see Appendix A for a more detailed discussion). It provides an alternative to both parametric and non-parametric estimation methods. Parametric methods include maximum-likelihood approaches and Bayesian approaches, and these approaches can be used to derive optimal methods for specific distributional families. Essentially all statistical modeling in vision makes use of parametric methods (see, e.g., Knill & Richards, 1996; Landy, Maloney, Johnston, & Young, 1995).

They are called "parametric" because they are based on the assumption that all statistical distributions are known except for a small number of parameters. A typical parametric problem is to estimate the mean μ and variance σ^2 of a Gaussian distribution. We can derive estimators that are unbiased and have minimum variance. If we were certain that the distribution of data is Gaussian it would be difficult to justify not using such estimators.

Non-parametric estimators, in contrast, are based on only weak distributional assumptions and typically have higher variance than corresponding parametric estimators. For any statistical problem, we can typically derive parametric estimators that have lower variance and methods for testing hypotheses that have greater statistical power than if we used non-parametric methods. However, if we assume that the distribution from which data is drawn is, for example, Gaussian and it is not, then our "optimal" parametric estimators need not be very "good" at all (Tukey, 1960).

Robust estimation provides an alternative to parametric and non-parametric approaches. The experimenter may know that the distribution is "close" to Gaussian but not precisely Gaussian. The distribution may be a mixture of two distributions, one of which is Gaussian and one of which is not. On some proportion ε of trials, the data are drawn from the non-Gaussian distribution, and this contamination may manifest itself by the presence of evident outliers in the data.

Robust statistics provides principled methods for dealing with these outliers—data points that are likely to have been generated by a contaminating process that is distinct

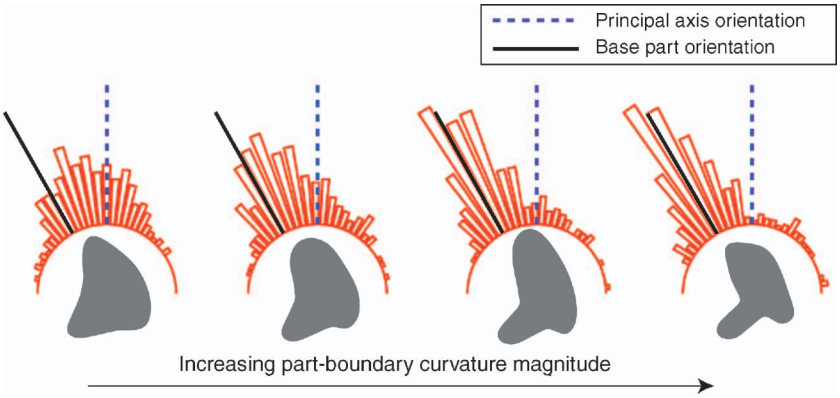

Figure 2. Representative stimuli and results from Cohen and Singh (2006). Observers adjusted an orientation probe to match the perceived orientation of each shape. Results are shown as polar histograms of pooled observer settings corresponding to four part-boundary conditions. Mean orientation settings for shapes with weak part boundaries were close to that of the global principal axis of the shape. However, with increasing sharpness of the part boundaries, mean orientation settings approached the orientation of the large base part.

from the presumed underlying model for the bulk of the data.

We might expect that an optimal statistical estimator for the Gaussian distribution would be "nearly optimal" when applied to data from distributions that are "nearly Gaussian." Surprisingly, it need not be. Tukey (1960) demonstrated that optimal estimators may fail dramatically with even small proportions of contamination ε. Robust statistics consists of methods that allow us to derive estimators that are close to optimal for an uncontaminated distribution but that are resistant to small amounts of contamination ε (Hampel, 1974; Huber, 1981). The "goodness" of each possible estimator is evaluated not only for the base distribution $f(x)$ but also for a neighborhood of contaminated distributions $(1 - \varepsilon)f(x) + \varepsilon h(x)$, where $h(x)$ is any distribution and $0 < \varepsilon \leq \varepsilon_0$. A typical choice of robust statistic is the one with the best "worst case" performance in this neighborhood (Huber, 1981).

To give a concrete example, if a sample is drawn from a Gaussian distribution with unknown mean μ, the mean of a data sample $\bar{X}$ is the unbiased minimum variance estimator of the parameter μ (Lehman, 1983, p. 84). If up to 1–2% of the data could be extreme values drawn from a second, unknown distribution then a 10% trimmed mean (where we discard the smallest 5% of the sample and the largest 5% of the sample) would typically be a lower variance estimator than the mean. If the outliers resulting from contamination are much larger in absolute value than the typical samples form the Gaussian, the 10% trimmed mean could be much lower in variance than the ordinary mean. But we can do even better (in terms of lower variance) than the 10% trimmed mean if we use a Huber-M estimator that assigns different weights to points depending on how extreme they are (Maronna, Martin, & Yohai, 2006, pp. 25–29). We will illustrate such an estimator in the General discussion section.

While robust methods deal with a broader class of contaminations than just the introduction of outliers, they are best known as principled methods for dealing with "suspicious" data points. Rather than having to make a binary choice between including or excluding a particular data point, a robust statistical estimator assigns differential weights to data points based on some quantity that reflects the likelihood that they arose from the same generative process as the bulk of the data or are instead due to the second contaminating process (Hampel, 1974; Huber, 1981; Hon et al., 1997; Landy et al., 1995). Data points likely to have arisen from the same generative process are assigned a full weight of 1, whereas those likely to have arisen from a different process are assigned weights near 0, with a continuous fall-off in between.

We examine whether the human visual system adopts a similar strategy in estimating a global property—the orientation of dot clusters—in the face of uncertainty concerning segmentation.

Experiment

We investigate the perceived orientation of dot clusters that could potentially be segmented into two sub-clusters. Our basic question is: How does the perceptual segmentation of a dot cluster influence its perceived orientation? In other words, as a cluster gradually goes from being perceived as "clearly a single cluster" to "probably two clusters" to "clearly two distinct clusters," how are observers' orientation estimates affected?

Consider the stimulus in Figure 1a, which could be interpreted either as a single cluster or as a composite consisting of a large elliptical sub-cluster and a much smaller sub-cluster. If the stimulus is treated as a single dot cluster, its orientation should be predicted by the principal axis of the entire cluster as a whole (shown by the red line in Figure 1b). Conversely, if the cluster is fully segmented, with the smaller sub-cluster treated as entirely separate, one would expect the dots within the small sub-cluster to be excluded entirely from the computation of orientation. Hence, the perceived orientation would be predicted by the principal axis of the larger cluster alone (shown by the blue line in Figure 1b).

These two strategies correspond to the opposite ends of a prediction spectrum, plotted in Figure 3 in terms of the "influence" of the smaller sub-cluster on the

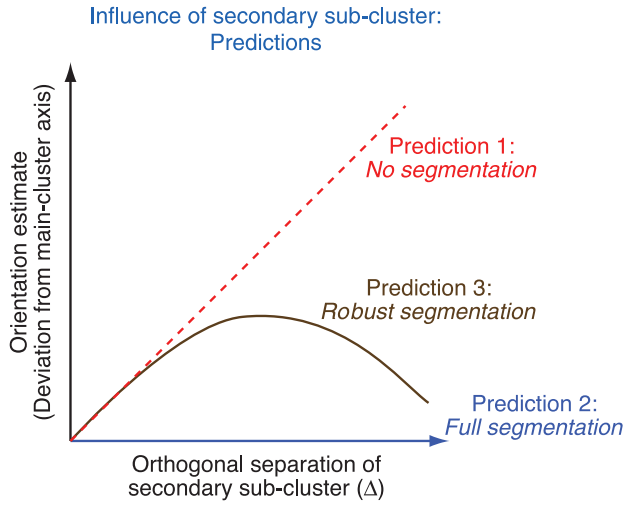

Figure 3. Influence of the secondary cluster on the estimated orientation of the dot clusters as a function of its orthogonal separation from the axis of the main cluster. Three predictions are shown based on three possible strategies. (1) *No segmentation:* All dots are treated equally, including the ones in the small sub-cluster. The influence continues to rise without bound. (2) *Full segmentation:* The secondary cluster is ignored entirely and hence exerts no influence whatsoever. (3) *Robust segmentation:* For small separations, the smaller sub-cluster is included entirely in the principal-axis computation (full influence) but, with increasing separation, its points are systematically down-weighted (decaying influence).

overall perceived orientation. According to the "full-segmentation" prediction (horizontal blue line), the smaller sub-cluster has no influence at all since it is excluded from the orientation computation. According to the "no segmentation" prediction (dashed red line), the smaller cluster is included fully in the orientation computation, i.e., treated equally with the dots in the large cluster. Hence, as the smaller cluster is moved further and further away from the main cluster, its influence on the orientation estimate continues to rise without bound.

The "robust segmentation" prediction is intermediate: For small separations between the two sub-clusters, the dots within the secondary sub-cluster are included in the orientation computation. However, with increasing separation (hence increasing likelihood of segmentation), these dots are systematically down-weighted. Thus, as the separation of the secondary sub-cluster increases, its influence on the overall orientation estimate at first rises (following the "no segmentation" prediction); but with further increase in separation, it gradually falls to zero, corresponding to the "full-segmentation" prediction (see Figure 3, solid brown curve). We ask whether the visual system employs a robust segmentation-like strategy in computing the orientation of dot clusters.

It is evident that the spatial separation between the two sub-clusters is not the only variable that affects their perceived segmentation. Given a fixed separation, sub-clusters with higher dot density are more likely to be perceived as "separate" or segmented than those with lower dot density. This is analogous to the simple fact in statistics that, for a given difference between two sample means, it is easier to find reliable evidence in favor of a two-distribution hypothesis over a one-distribution "null" hypothesis, with a larger sample size.[3] In terms of the influence function, one would thus expect that with increasing separation between the two sub-clusters, the influence function for high-density dot clusters would peak and fall sooner (i.e., at smaller spatial separations) than that for low-density dot clusters—because the high-density sub-clusters presumably require a smaller spatial separation to be reliably segmented. Our experiment manipulates both the separation between the two sub-clusters and their (common) dot density.

Methods

Observers

Five observers at Rutgers University, with normal or corrected-to-normal visual acuity, participated in the experiment. All were naïve to the purpose of the experiment.

Stimuli and design

Stimuli consisted of dot clusters formed by the union of two sub-clusters: (i) a large sub-cluster comprising dots sampled uniformly from an elliptical region and (ii) a secondary sub-cluster comprising dots sampled uniformly from a small circular region. The dimensions of the ellipse were 10.9 dva by 4.36 dva. The smaller cluster was placed such that the center of the circular region lay on an (invisible) line orthogonal to the major axis of the ellipse, placed 3.27 dva from the center of the large ellipse (see Figure 4). Each dot had a diameter of 5.2 arcmin.

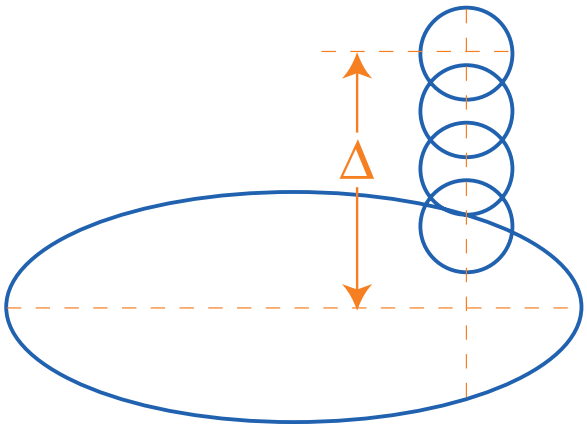

Figure 4. A schematic diagram showing the underlying ellipse and disks used to generate the dot stimuli. The small circular region was placed at one of four possible distances Δ from the main axis of the large ellipse.

The variables manipulated were as follows:

1. The separation Δ between the two sub-clusters, as measured by the orthogonal distance of the center of the small circle from the major axis of the ellipse. The four values of Δ used were 1.52, 2.61, 3.7, and 4.79 dva.
2. The mean density of dots in the sub-clusters. The three values used were 0.63, 1.26, and 1.89 dots per square dva. These density values determined the number of dots to be sampled within the elliptical and circular regions on any given trial.
3. The diameter of the small circular region: 1.74 and 2.4 dva.

There were thus a total of $4 \times 3 \times 2 = 24$ conditions. Figure 5 shows examples of stimuli in eight of these conditions: dot clusters in the lowest and highest density conditions, each with the four levels of spatial separation between the two sub-clusters (for the smaller sized secondary sub-cluster).

Procedure

On each trial, an observer was shown the test dot cluster composed of light blue dots against a black background for 1000 ms, followed by a mask composed of interleaved white and light blue dots for 250 ms. An adjustable probe pattern then appeared on the screen, consisting of multiple

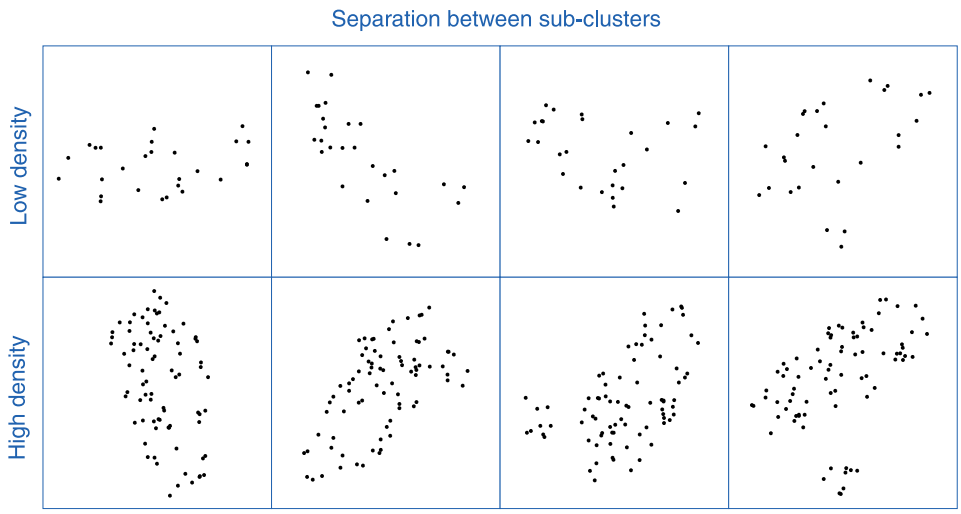

Figure 5. Examples of dot cluster stimuli shown for 8 of the 24 conditions used in the experiment. The top row shows clusters with the lowest dot density (0.63 dots per dva^2), the bottom row with the highest density (1.89 dots per dva^2). The four columns correspond to the four levels of spatial separation between the center of the small circular region and the major axis of the elliptical region (Δ = 1.52, 2.61, 3.7, and 4.79 dva). All stimuli shown here have the smaller diameter used for the secondary cluster (1.74 dva).

parallel red lines (see Figure 6). The observers' task was to adjust the orientation of the probe pattern using a track ball. They were instructed as follows: "Rotate the red probe pattern to match the perceived orientation of the dot cluster. If you see multiple possible orientations, choose the strongest one. There is no right or wrong answer." No mention was made in the instructions about the possibility of a secondary sub-cluster, and observers relied on their own perceptual intuitions to interpret the notion of "test dot cluster." The multi-line pattern was used instead of a single line segment because we were interested in the perceived orientation of the dot cluster regardless of the specific point within the cluster through which the principal axis should pass (e.g., its perceived center). No time constraints were imposed on the adjustment.

Observers performed adjustments in 40 trials for each of the 24 conditions. The trials within each condition were counterbalanced for handedness, i.e., the position of the secondary sub-cluster relative to the main elliptical sub-cluster. For each trial, a new dot configuration was sampled with the parameter values determined by the required condition. The clusters were shown at a random

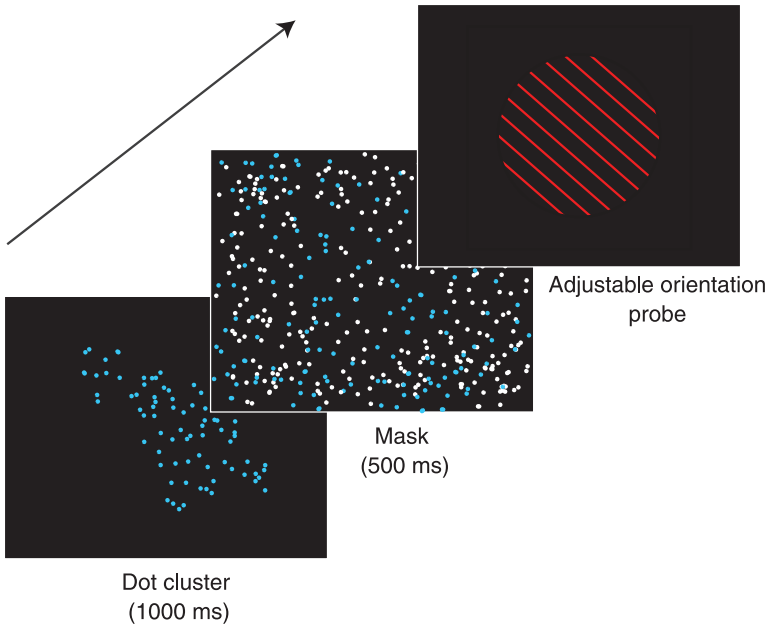

Figure 6. The trial sequence used in the experiment. Observers adjusted the orientation of the multi-line probe to match the perceived orientation of the test dot cluster.

orientation. Each observer's settings were collected in two experimental sessions and were preceded by a practice session.

Results

Observers' orientation settings were encoded in terms of the angular deviation from the major axis of the underlying ellipse used to generate the main sub-cluster. Deviations in the direction of the principal axis of the entire cluster (i.e., of the union of the two sub-clusters) were considered positive. Preliminary analyses of the orientation settings revealed only small differences between corresponding left-handed and right-handed stimuli, and their directions were inconsistent across observers (mean differences = −3.03, −2.3, 3.18, 3.54, 3.98 degrees for the five observers). For subsequent analyses, we collapsed the data across left-handed and right-handed shapes and focused on the effects of spatial separation, dot density, and size of the smaller sub-cluster. Figure 7 plots the mean orientation settings in degrees as a function of the three independent variables.

The *x*-axis in each plot thus corresponds to the prediction of the "full-segmentation" hypothesis discussed above (recall Figure 3), namely, the orientation settings that would be predicted if observers treated the smaller sub-cluster as completely separate, and ignored it entirely in computing the principal axis of the dot cluster. The dashed lines in each plot depict the predictions of the "no-segmentation" hypothesis. The three dashed lines correspond to the three different levels of dot density and were computed based on the actual (sampled) dot configurations shown during the course of the experiment. These are thus the mean orientation settings that would be predicted if observers treated each cluster as a single, unsegmented, perceptual unit and computed its principal axis by weighting all dot equally.

As is clear from the plots in Figure 7, observers' settings follow neither of these two predictions but rather exhibit a pattern that is qualitatively similar to the "robust segmentation" hypothesis discussed above. For small values of spatial separation Δ between the sub-clusters, observers' settings follow the "no segmentation" prediction (the diagonal dashed lines), indicating that all dots are being treated equally in computing orientation. With further increases in spatial separation, however, the orientation settings gradually fall and approach the "full-segmentation" prediction (the *x*-axis of each plot)—indicating that the dots within the small cluster are ignored almost entirely in these conditions.

Differentially weighted principal axis computation

In order to quantify the influence of the dots within the secondary sub-cluster, we analyzed the data in terms of a

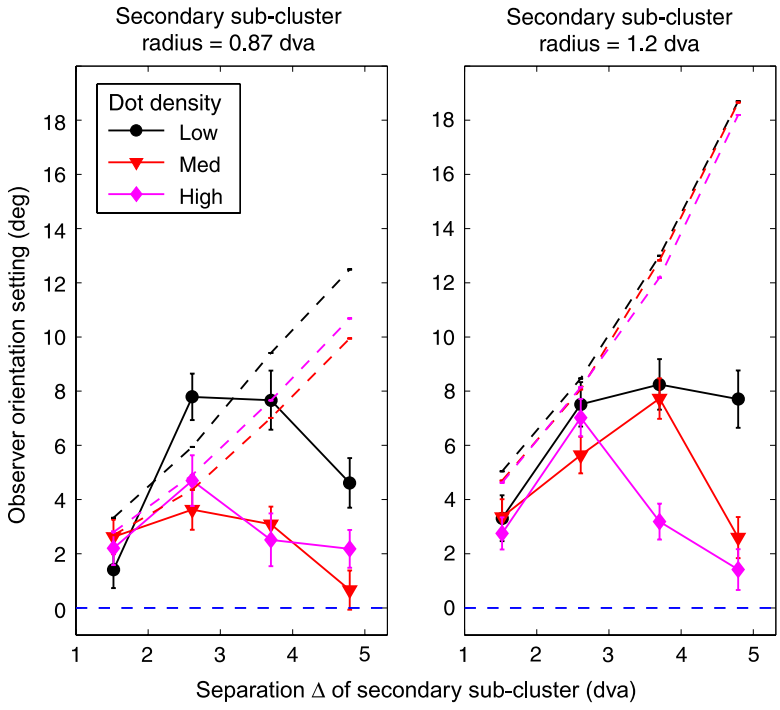

Figure 7. Mean orientation settings measured as angular deviations from the major axis of the ellipse used to generate the large cluster. The two plots correspond to the two different sizes of the secondary cluster. The three curves in each plot correspond to the three different dot densities: 0.63, 1.26, and 1.89 dots per dva^2. The three oblique dashed lines correspond to the predictions based on the principal axis of the entire cluster (i.e., with the dots within the secondary cluster fully included in the computation).

differentially weighted principal-axis computation: The dots within the main sub-cluster were assigned a "full" weight of 1, whereas dots within the smaller sub-cluster were assigned a partial weight of α. For each trial, we determined the value of the partial weight α (to the dots within the smaller sub-cluster) that would yield the observed orientation setting on that trial. We did this by computing the orientation setting as a function of α numerically for many finely spaced values of α and selecting the value of α that led to (or best approximated) the observed setting on that trial. These estimated values of the weights α were then averaged within each given condition. Figure 8 plots these mean weights as a function of spatial separation, dot density, and secondary-cluster size. A value of 0 along the *y*-axis in Figure 8 corresponds to the prediction of the "full-segmentation" hypothesis, whereas a value of 1 corresponds to the prediction of the "no-segmentation" hypothesis, i.e., the principal axis of the entire cluster.

As before, the plots in Figure 8 make it clear that for smaller values of spatial separation (Δ = 1.52 and 2.61 dva), observers' orientation settings follow the prediction of the "no-segmentation" hypothesis; the values of the weights α assigned to dots within the smaller cluster are statistically indistinguishable from 1. But with further increase in spatial separation (Δ = 3.7 and 4.79 dva), the orientation settings approach the predictions of the "full-segmentation" prediction (i.e., the partial weights α approach 0).

The plots in Figure 8 also clarify the contribution of dot density to perceived orientation. The partial weights α for the highest dot density (magenta curves) are consistently lower—i.e., closer to the "full-segmentation" prediction—than those for the lowest dot density (black curves). This is consistent with the prediction that, given a particular level of spatial separation between the two sub-clusters, dot clusters with higher dot density are more likely to be perceived as segmented. (In other words, dot clusters with higher dot density require a smaller spatial separation in order to be reliably segmented.)

Expressed as partial weights α in a differentially weighted principal-axis computation, the data no longer show an effect of size of the secondary sub-cluster. Therefore, the difference between the left and the right subplots in Figure 7 can be attributed entirely to the fact that a larger secondary sub-cluster necessarily exerts a greater influence on the principal axis of the overall cluster than a smaller one.

The results are thus consistent with a differentially weighted principal-axis computation in which the dots within the smaller sub-cluster are assigned systematically lower weights (i) with increasing spatial separation between the sub-clusters and (ii) with increasing dot density. Given that increasing spatial separation and increasing dot density both result in an increased likelihood of segmentation, these results are qualitatively consistent with a robust estimation approach to the computation of perceived orientation.

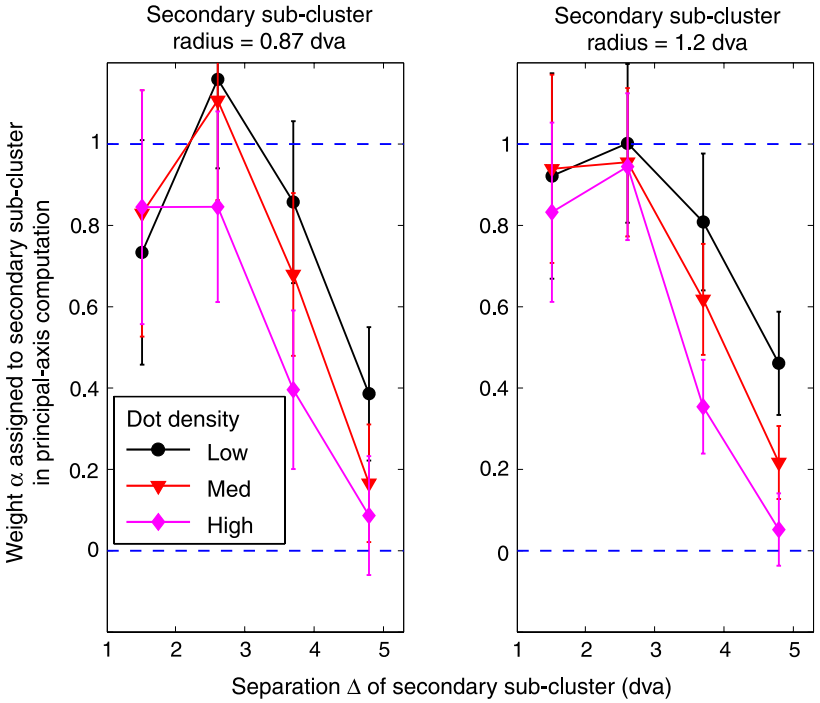

Figure 8. Orientation settings recoded in terms of the relative weight α assigned to the secondary cluster dots in a differentially weighted principal-axis computation that would yield the observer's setting on any given trial. A value of 1 along the *y*-axis corresponds to the prediction of the "no segmentation" hypothesis, whereas a value of 0 corresponds to the prediction of the "full segmentation" hypothesis.

General discussion

The estimation of any visual property requires that information be pooled over an appropriately defined region and thus depends critically on perceptual segmentation. We have suggested that (i) perceptual segmentation is better treated as a graded process, and (ii) the visual estimation of a global visual property (such as orientation) can be used as a tool to measure probabilistic segmentation, i.e., the strength of the visual system's "belief" concerning whether an image region corresponds to one perceptual unit or two.

We used the context of perceived orientation of dot clusters formed by the union of two sub-clusters: a large elliptical sub-cluster and a small circular sub-cluster. By manipulating the spatial separation between the two sub-clusters and their common dot density, we varied the strength of evidence in favor of the "segmentation" hypothesis—that the smaller sub-cluster is a separate perceptual unit from the larger one. The results showed a systematic dependence of perceived orientation on both variables. For small spatial separations between the two sub-clusters, observers' orientation estimates were consistent with the principal axis of the entire cluster (indicating that the dots within the smaller sub-cluster were being treated equally with the rest). However, with increasing spatial separation—hence increasing evidence for segmentation—the orientation estimates gradually approached the principal axis of the larger sub-cluster alone (indicating that the dots within the smaller cluster were having little influence). Moreover, this decrease in influence occurred sooner, i.e., for smaller spatial separations, when the dot density was high. This is consistent with the fact that, given a fixed spatial separation, a higher dot density provides more reliable evidence for segmentation. The above pattern of results is qualitatively consistent with a robust strategy for estimating orientation.

Robust principal components analysis

The procedure just described makes explicit use of knowledge of the cluster membership of each dot. In effect, it assumes that it has available a visual parsing of the scene into clusters (objects) that is correct or close to correct. Typical robust methods in statistics do not have access to such information, and we decided to compare the performance of a typical robust PCA algorithm based on an iterative weighted principal components analysis.

This PCA algorithm is very similar in structure to common robust estimation approaches that use Huber-M estimators (Maronna et al., 2006, pp. 25–29). The algorithm was iterative, computing a series of estimates of the principal axis, continuing until the estimates converged.

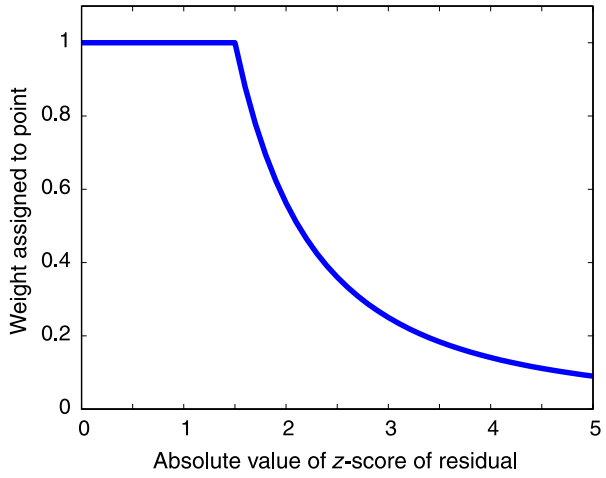

Figure 9. The Huber-*k* type weighting function used in the robust PCA analysis. Points with *z*-scored residuals in the interval $[-k, k]$ are assigned a full weight of 1. Those with larger residual magnitudes are assigned weights given by k/z^2.

It began by computing the principal axis of all dots in the stimulus array including those in both the main and the secondary cluster using ordinary (non-robust) PCA. Then it computed the signed residual differences between each point in the array and the nearest point on the principal axis and converted these signed residuals to *z*-scores.

These *z*-scores were then converted to weights by a weighting function (the Huber-*k* loss function shown in Figure 9). The weighting function assigns full weight to *z*-scores between $-k$ and $+k$ and a lower score to more extreme *z*-scores, decreasing with absolute magnitude of the *z*-score.[4] We selected the value $k = 1.5$ for this example, but performance is not sensitive to k. The principal axis was then recomputed by PCA but now weighting each point according to the weights assigned by the Huber-*k* loss function.

These two steps (estimation of the principal axis, computation of residuals and weights) were then repeated until the estimate of the principal axis converged. The final weights assigned to the points serve as a measure of how close each point was to the principal axis and how much weight it had in determining the final principal axis. At each iteration in the algorithm, the effect of any point on the next principal axis estimate depended only on its residual with respect to the previous principal axis estimate.

There are competing robust PCA approaches (Maronna et al., 2006, Section 6.1); we chose this approach because it highlights possible qualitative differences between human performance and statistical approaches that do not have a visual parsing of the scene.

In Figure 10, we plot an influence function for the effect of smoothly displacing the secondary cluster away from the main. The horizontal axis is used to plot the displacement Δ of the secondary cluster. When Δ was 0, the secondary cluster was contained in the main cluster. The

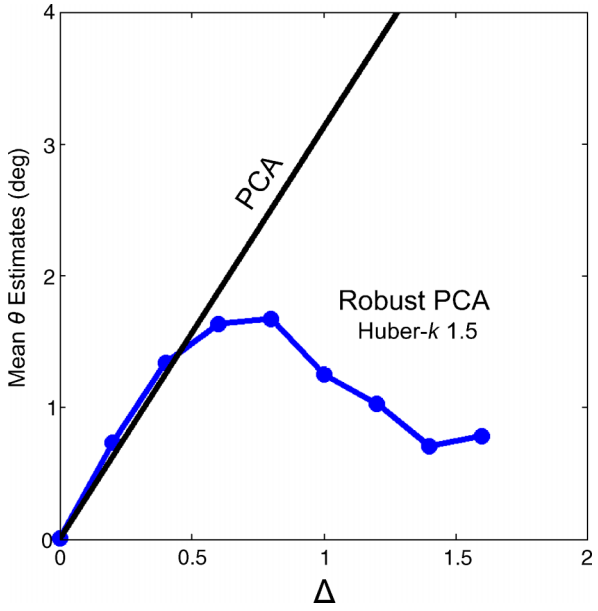

Figure 10. The results of the Robust PCA analysis showing the influence of the secondary cluster dots as a function of their increasing spatial separation Δ from the large cluster. Each point corresponds to the mean of 300 fits of simulated stimuli with a given value of Δ. The results of using ordinary (non-robust) PCA are shown for comparison.

vertical axis plots the angle of the estimated principal axis based on weighted principal components analysis. The results are averaged across 300 fits of simulated stimuli for each value of Δ. The influence of the secondary cluster is evident for small values of Δ but, for large values, the influence diminishes, characteristic of a robust estimator. For comparison, we show the results of using ordinary (non-robust) PCA.

The robust PCA method just described has several drawbacks that call into question its value as model of human performance. First, the loss functions are defined on the standardized values of residuals (measured from the current estimate of principal axis in the iterative procedure). This means that when points are down-weighted based on the likelihood that they arose from a separate generative process, this determination is made simply based on their standardized residuals independent of whether any particular group of deviant points form a spatially coherent cluster.

As a consequence, an increase in the dot density of both sub-clusters results in little change in the assignment of weights or the influence function. In a simulation, we found that tripling the dot density had almost no effect on the influence function and on where it peaked. Again, the performance of the algorithm is at odds with observed behavior of our observers.

Moreover, suppose that there are two secondary clusters, both well outside of the main cluster but one much closer to it than the other. The residuals to the more deviant cluster disproportionately affect the standardized residual scale and, as a result, the points in the other deviant sub-cluster will be assigned z-scores near 0 and weights near 1 and will exert considerable influence on the final estimate of the principal axis. It seems plausible that the visual system would segment both the further cluster and the closer, ignoring both, but a residual-based algorithm cannot do so easily.

The above failures stem from the fact that a robust estimation strategy based on residuals embodies an overly simplistic generative model of "objects." Our results suggest—but do not conclusively demonstrate—that the visual ability to detect clusters in scene plays a crucial role in segmentation and estimation of object orientation in human vision. We advance this claim as a conjecture.

To our knowledge, no existing robust statistic makes use of a similar model-based approach and, if our conclusions above are valid, there is no point in comparing human performance to existing robust estimators. Even if a standard robust estimator duplicates or exceeds human performance, it is not using the same part-based approach: it is not a model of human visual processing in our experiments.

An interesting direction for future work would be to investigate psychologically valid generative models of dot clusters. A long and fruitful line of research on shape representation has focused on axis or skeleton-based models of objects (Blum, 1973; Blum & Nagel, 1978; Brady & Asada, 1984; Leyton, 1987). It is likely that probabilistic versions of skeletal models (e.g., Feldman & Singh, 2006; Zhu, 1999)—suitably adapted to dot clusters—would provide a promising candidate. The investigation of such models and their appropriateness for modeling the visual detection and segmentation of dot clusters awaits future research.

Appendix A

Robust statistics is a collection of methods for deriving good estimators when the experimenter has only partial knowledge of the distribution from which data was drawn. These methods can, in principle, be used to select optimal statistics by any criterion. For simplicity, we focus on estimators that have low variance and that are unbiased.

Suppose, for example, that the experimenter has a sample of data drawn from a Gaussian distribution $f(x;\ \mu,\ \sigma^2)$,

$$X_1, \ldots, X_n \sim f\left(x; \mu, \sigma^2\right), \tag{A1}$$

with unknown mean μ and unknown variance σ^2. Suppose that the experimenter's goal is to estimate the unknown parameter μ, the mean of the distribution. Given this

information, we can derive the estimator $T(X_1,\ldots, X_n)$ that is unbiased (the expected value of the estimator $E[T(X_1,\ldots, X_n)] = \mu$) and that has the smallest variance among all unbiased estimators.

In the Gaussian case, the unbiased estimator with the smallest variance is the mean of the data, $T(X_1,\ldots, X_n) = \bar{X}$. Suppose that, for example, if we take the mean of a sample of size 100 from a Gaussian distribution with unknown mean and variance 1 and compare the mean with another unbiased statistic, the median of the sample, the median is also unbiased but its variance is about 60% greater than that of the mean.

If we change the distribution to, for example, Student's *t*-distribution with 3 degrees of freedom, the roles of the median and the mean reverse, and now the median has lower variance than the mean by about 40%.

Thus, if the distribution is not Gaussian, the sample mean may no longer be the unbiased estimator with minimum variance, or it may not be an unbiased estimate of the population mean.

But what if the statistician is unsure about the distributional family? Perhaps the sample is drawn from a Gaussian distribution with probability $1 - \varepsilon$ but with a small probability ε from a second unknown distribution, $h(x)$. The resulting distribution is a mixture distribution:

$$X_1, \ldots, X_n \sim (1 - \varepsilon)\phi\left(x; \mu, \sigma^2\right) + \varepsilon h(x). \quad \text{(A2)}$$

We still wish to estimate μ, σ^2 but now we have the added complication that part of our data may be drawn from an unknown second distribution $h(x)$.

Tukey (1960) demonstrated that estimators that are optimal in the uncontaminated case may behave very badly with even very small degrees of contamination ε (Tukey, 1960). They are not "robust" to failures of the distributional assumptions.

When the true distribution is unknown, the statistician may choose to use non-parametric methods that make only weak assumptions about distribution. The median of a sample is an estimate of the median of the distribution for any distribution. Since the mean and the median of the Gaussian are both μ, the mean and the median are in competition as estimators. In the uncontaminated case, for large samples, the median has a variance that is more than 60% larger than that of mean. However, with even small amount of contamination, the median can have lower variance than the mean.

The goal of robust statistics is to provide a principled, third alternative by looking for statistics that do almost as well as the optimal statistic in the parametric case but which are little affected by "small" amounts of contamination ε. A more precise criterion might be to rate each possible statistical estimator by its worst performance for any choice of the unknown contamination $h(x)$.

One measure of use in evaluating the robustness of a statistic is its *influence function* (Hampel, 1974). We restrict attention to contaminated distributions that are Dirac (impulse) functions at location x,

$$f_\varepsilon(x) = (1 - \varepsilon)f(x) + \varepsilon\delta(x - x_0), \quad \text{(A3)}$$

where $\delta(x)$ denotes a Dirac generalized function (Bracewell, 2000, chap. 5). The contaminated distribution above is readily interpreted. With probability ε, the value x is substituted for the sample value that would have been drawn from $f(x)$. We denote any estimator by $T_n(X_1,\ldots, X_n)$. If we increase the sample size n, we can compute the limit of T_n denoted T_∞, and we can write $T_\infty(f(x))$ as shorthand for the limit as sample size increases when samples are drawn from the distribution $f(x)$. If, for example, $f(x)$ is Gaussian with mean μ, and $T_n(X_1,\ldots, X_n) = \bar{X}$, then $T_\infty(f(x)) = \mu$: the limit of the mean of the sample converges to the population mean as sample size increases.

The influence function is a Gâteau derivative (Jurečková & Picek, 2005, p. 15)

$$T_x'(f) = \lim_{\varepsilon\to 0+} \frac{T_\infty(f_\varepsilon(x)) - T_\infty(f(x))}{\varepsilon}, \quad \text{(A4)}$$

which measures how much effect a small contamination has on the statistic $T_n(X_1,\ldots, X_n)$. The influence function we report in the text is an empirical version of this theoretical limit.

A robust estimator $T_n(X_1,\ldots, X_n)$ has an influence function that goes to 0 as $|x|$ increases or is at least bounded. In contrast, we can show that for the ordinary mean $T_n(X_1,\ldots, X_n) = \bar{X}$, the influence function $T_x'(f) = x$ increases without bound. The mean is not a robust estimator.

Acknowledgments

This research was funded in part by Grants BCS-0216944 and CCF-0541185 from the National Science Foundation (MS) and Grant EY08266 from the National Institute of Health (LTM). We thank Jacob Feldman, Mordechai Juni, and Eileen Kowler for helpful discussions and M. Juni for his help with running the experiment.

Commercial relationships: none.
Corresponding author: Manish Singh.
Email: manish@ruccs.rutgers.edu.
Address: Department of Psychology, Rutgers University, New Brunswick/Piscataway Campus, 152 Frelinghuysen Road, Piscataway, NJ 08854-8020, USA.

Footnotes

[1]More generally, the first principal axis is given by the eigenvector of the covariance matrix corresponding to the largest eigenvalue. Throughout we will simply use "principal axis" to refer to the first principal axis.

[2]Blum's (1973) medial-axis transform treats each shape as the union of maximally inscribed disks. The locus of the centers of these circles defines the medial axis, which provides a skeletal description of the shape.

[3]Surprisingly, however, subjects appear not to be very good at taking into account sample size in making intuitive cognitive judgments concerning whether or not two samples are drawn from the same population (see Obrecht, Chapman, & Gelman, 2007).

[4]This robust estimator is an example of a common method adopted from robust regression. Essentially any curve that smoothly decays to 0 would lead to better performance than a non-robust estimator. The resulting estimator would be robust to some extent but would not have the qualitative properties that human observers exhibit. The choice of an "optimal" robust estimator depends on not just the assumed distribution of that data but possible small deviations from that distribution and typically there is no one correct choice.

References

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

Blum, H. (1973). Biological shape and visual science: I. *Journal of Theoretical Biology, 38,* 205–287. [PubMed]

Blum, H., & Nagel, R. N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition, 10,* 167–180.

Bracewell, R. N. (2000). *The Fourier transformation and its applications* (3rd ed.). New York: McGraw-Hill.

Brady, M., & Asada, H. (1984). Smoothed local symmetries and their implementation. *International Journal of Robotics Research, 3,* 36–61.

Burbeck, C. A., & Zauberman, G. S. (1997). Across-object relationships in perceived object orientation. *Vision Research, 37,* 879–884. [PubMed]

Cohen, E. H., Schnitzer, B. S., Gersch, T. M., Singh, M., & Kowler, E. (2007). The relationship between spatial pooling and attention in saccadic and perceptual tasks. *Vision Research, 47,* 1907–1923. [PubMed]

Cohen, E. H., & Singh, M. (2006). Perceived orientation of complex shape reflects graded part decomposition. *Journal of Vision, 6*(8):4, 805–821, http://journalofvision.org/6/8/4/, doi:10.1167/6.8.4. [PubMed] [Article]

Denisova, K., Singh, M., & Kowler, E. (2006). The role of part structure in the perceptual localization of a shape. *Perception, 35,* 1073–1087. [PubMed]

De Winter, J., & Wagemans, J. (2006). Segmentation of object outlines into parts: A large scale integrative study. *Cognition, 99,* 275–325. [PubMed]

Driver, J., & Baylis, G. C. (1996). Edge assignment and figure-ground segmentation in visual short-term matching. *Cognitive Psychology, 31,* 248–306. [PubMed]

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision, 2*(4):5, 324–353, http://journalofvision.org/2/4/5/, doi:10.1167/2.4.5. [PubMed] [Article]

Feldman, J. (1997). Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research, 37,* 2835–2848. [PubMed]

Feldman, J. (2001). Bayesian contour integration. *Perception & Psychophysics, 63,* 1171–1182. [PubMed]

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. [PubMed] [Article]

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41,* 711–724. [PubMed]

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69,* 383–393.

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

Hoffman, D. D., & Singh, M. (1997). Salience of visual parts. *Cognition, 63,* 29–78. [PubMed]

Hon, A., Maloney, L. T., & Landy, M. S. (1997). The influence function for visual interpolation. *SPIE Proceedings Series, 3016,* 409–419.

Huber, P. J. (1981). *Robust statistics*. New York: Wiley-Interscience.

Jurečková, J., & Pickek, J. (2005). *Robust statistical methods in R*. New York: Chapman & Hall.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.

Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt Brace.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative gestalt theory. *Psychological Science, 6,* 225–234.

Landy, M. S., Maloney, L. T., Johnston, E., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35,* 389–412. [PubMed]

Lánský, P., Yakimoff, N., & Radil, T. (1987). On visual orientation of dot patterns. *Biological Cybernetics, 56,* 389–396. [PubMed]

Lánský, P., Yakimoff, N., Radil, T., & Mitrani, L. (1989). Errors in estimating the orientation of dot patterns. *Perception, 18,* 237–242. [PubMed]

Lehman, E. L. (1983). *Theory of point estimation*. New York: Wiley.

Leyton, M. (1987). Symmetry-curvature duality. *Computer Vision, Graphics, and Image Processing, 38,* 327–341.

Li, W., & Westheimer, G. (1997). Human discrimination of the implicit orientation of simple symmetrical patterns. *Vision Research, 37,* 565–572. [PubMed]

Liu, B., Dijkstra, T. M., & Oomes, A. H. (2002). The beholder's share in the perception of orientation of 2-D shapes. *Perception & Psychophysics, 64,* 1227–1247. [PubMed]

Maronna, R. D., Martin, R. A., & Yohai, V. J. (2006). *Robust statistics, theory and methods*. Chichester, UK: Wiley.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biological Sciences, 200,* 269–294. [PubMed]

Melcher, D., & Kowler, E. (1999). Shape, surfaces and saccades. *Vision Research, 39,* 2929–2946. [PubMed]

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://journalofvision.org/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. [PubMed]

Obrecht, N., Chapman, G. B., & Gelman, R, (2007). Intuitive *t* tests: Lay use of statistical information. *Psychonomic Bulletin & Review, 14,* 1147–1152. [PubMed]

Oomes, A. H., & Dijkstra, T. M. (2002). Object pose: Perceiving 3-D shape as sticks and slabs. *Perception & Psychophysics, 64,* 507–520. [PubMed]

Peterson, M., & Gibson, B. (1994). Must figure-ground perception precede object recognition? An assumption in peril. *Psychological Science, 5,* 253–259.

Rock, I., & Brosgole, L. (1964). Grouping based on phenomenal proximity. *Journal of Experimental Psychology, 67,* 531–538. [PubMed]

Rubin, E. (1958). Figure and ground. In D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 194–203). Princeton, NJ: Van Nostrand. (Original work published 1915).

Singh, M., & Fulvio, J. M. (2005). Visual extrapolation of contour geometry. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 939–944. [PubMed] [Article]

Singh, M., & Hoffman, D. (2001). Part-based representations of visual shape and implications for visual cognition. In T. Shipley & P. Kellman (Eds.), *From fragments to objects: Grouping and segmentation in vision. Advances in psychology series* (vol. 130, pp. 401–459). New York: Elsevier Science.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin (Ed.), *Contributions to probability and statistics* (pp. 448–485). Stanford University Press.

Vishwanath, D., Kowler, E., & Feldman, J. (2000). Saccadic localization of occluded targets. *Vision Research, 40,* 2797–2811. [PubMed]

Warren, P. A., Maloney, L. T., & Landy, M. S. (2002). Interpolating sampled contours in 3-D: Analyses of variability and bias. *Vision Research, 42,* 2431–2446. [PubMed]

Warren, P. A., Maloney, L. T., & Landy, M. S. (2004). Interpolating sampled contours in 3D: Perturbation analyses. *Vision Research, 44,* 815–832. [PubMed]

Wertheimer, M. (1923). Laws of organization in perceptual forms. *Psychologische Forschung, 4,* 301–350.

Yakimoff, N. (1981). Does assignment of orientation to dot patterns reveal optimization processes in the visual system? *Biological Cybernetics, 42,* 39–43. [PubMed]

Yodogawa, E. (1985). Quantitative measure of perceived orientation strength of dot patterns. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics,* 584–588.

Zhu, S. C. (1999). Embedding Gestalt laws in Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21,* 1170–1187.