Chapter 4

Perceptual Shape Metrics

4.1 Introduction

4.1.1. What is shape?

When asked this question during my qualifying exam, I was stumped. My rehearsed answer should have been that shape is an object property that remains invariant under certain transformations, such as translation, rotation, reflection, scale and some deformations. This is the standard mathematician's view of shape. But the real question is: how do we represent shape in a way that is perceptually meaningful?

The psychologist will argue that the definition of shape is much more subtle. There is a rich history of investigations into shape perception. Mach (1914) first noted that a square rotated by 45 degrees is perceived as a diamond, not as a rotated square, demonstrating the importance of reference frames. The most complete, yet rarely cited work is that of Goldmeier in his 1936 thesis, later published as a monograph in 1972. Goldmeier demonstrated that the laws of perceptual organization can override certain obvious physical attributes of the stimulus in determining the perceived degree of similarity between two figures. Refer to Palmer (1999) for a complete overview of qualitative shape perception.

In this chapter, we adopt a narrow view of shape, close to that of the mathematicians. We would like to know if there are shape metrics that can be used to measure human sensitivity to small perturbations of shape.

4.1.2. Quantitative Shape

When two shapes are *qualitatively* different, like triangles and quadrilaterals, it makes little sense to talk about how similar their shapes are. They may be very far apart in "shape space", much like red and green are distant in color space. However, once we know we are comparing two figures that are nearby in shape space (comparing apples with apples, so to speak), it may be possible to define a *quantitative* measure of shape similarity. Again, this can be thought of as analogous to the mapping of color space, in which the just-noticeable-differences (jnds) between nearby colors is well defined. The challenge is to find a metric for shape space that agrees with human perception.

4.2 Shape space

4.2.1. Shape of natural forms

Thompson (1961), Rashevsky (1948) and many others agree that we get most of our shape information from the contour of an object. As Attneave (1954) pointed out, contours have relatively high information content, especially as compared to interior shading, etc. Stemming directly from the work of Thompson (1961) and Bookstein

(1985; 1989), the method of deformable templates has evolved as one way to gauge the distance between two shapes. These methods rely on defining homologous or corresponding points between the two objects, although some methods manage to avoid this (e.g., Belongie, Malik & Puzicha, 2002).

Representing shape as a set of points is not unreasonable. Wagemans, De Winter and Panis (2002) extended Atteneave's (1954) observations in a large-scale study and demonstrated that we are quite good at recognizing objects from straight-lined drawings when the landmark points are preserved (Figure 4.1). Landmark points can include salient semantic features (corner of the eye, tip of the nose) or simple extremes of curvature.

To get a measure of shape *similarity* we would like to introduce a metric to account for the distance between the point clouds representing two shapes. There are several measures one might consider, two of which we will evaluate.

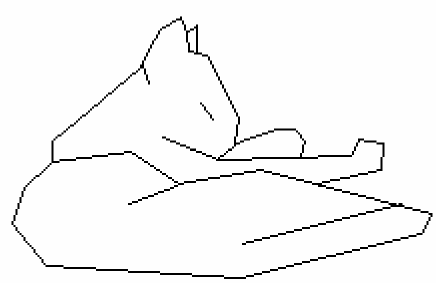

Figure 4.1. Attneave's sleeping cat. The sketch is made by drawing straight lines between extrema of curvature on the image of a sleeping cat.

4.2.2. Procrustes' distance

The first measure we consider is one of the simplest, Procrustes distance (*Pdist*). Pdist measures the sum of squares of Euclidean distances between known corresponding points, after the two sets have been aligned to correct for translation and scale.

$$Pdist = \sum_i \left\| z_i^A - z_i^B \right\|^2$$

where z is the complex number $z = x + iy$ for point sets A and B containing i points in two dimensions.

4.2.3. Kendall's shape space

Kendall (1984; 1989) described shape spaces for comparing point sets in any number of dimensions. While the space is directly related to Procrustes' distance, it undergoes a Riemannian submersion to create a non-Euclidean metric. Distances between point sets are then measured as geodesics along the surface of a hyper-sphere. We refer to these distances as Kendall's distance (*Kdist*). For point sets in three or more dimensions, the space contains singularities.

$$\cos^2\left(Kdist\right) = \frac{\left(\sum_i z_i^A \bar{z}_i^B\right)\left(\sum_i \bar{z}_i^A z_i^B\right)}{\left(\sum_i z_i^A \bar{z}_i^A\right)\left(\sum_i z_i^B \bar{z}_i^B\right)}$$

where $\bar{z}$ is the complex conjugate of z, $\bar{z} = x - iy$. Before comparing two point sets, they are corrected for translation by placing their centroids at zero by setting $\sum_i z_i = 0$.

4.3 Validating the metrics

> *"Of all the conceivable physical measures of shape, analytical as well as gestalt, there are undoubtedly many that have little or no value from a psychophysical point of view… Unfortunately, there is no quick and easy way to determine which physical measurements have greatest psychological relevance; only experimentation can answer this question."*
>
> *Attneave and Arnoult (1956)*

We examine *Kdist* and *Pdist* experimentally to see if either has greater psychological relevance to the judgment of shapes. We avoid object parts in the description of the shape by using only two dimensional convex polygons in these experiments. A polygon can be defined as a point cloud by its vertices.

If *Kdist* and/or *Pdist* are to be validated as a perceptual metric, they must first satisfy the criteria of producing a stable measure of our perceptual threshold for the just-noticeable-difference (jnd) of simple perturbations to a reference shape. We consider two perturbations: (1) random jitter of the vertices and (2) systematic affine transformation of the shape.

4.3.1. Experimental methods

Subjects

Four naïve university students participated in these experiments. None had ever participated in psychophysical experiments before, so some supervised training was required initially. Subjects participated for 9.5 to 25 hours depending on the number of shapes they studied. Refer to CPHS protocol 2000-3-3.

Stimuli

Random quadrilateral reference shapes were generated by placing four points on a two dimensional Cartesian grid. The reference shapes were then systematically altered by left-multiplying the vertex coordinates by

$$R(-\theta)\begin{bmatrix} 1+\Delta a & 0 \\ 0 & \dfrac{1}{1+\Delta a} \end{bmatrix} R(\theta)$$

where $R(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ and Δa is the amount of affine change to add to the shape. Δa varies from 0.01 to 0.15 in steps of 0.01 and θ is chosen randomly from a uniform distribution over $[0, \pi]$. We rotate the shape before applying the affine transformation so that the distortion can occur in any direction. Each vertex undergoes the same transformation.

To create randomly altered shapes, jitter is applied to both the x and y components of each vertex. The magnitude of the jitter for each direction is drawn from the uniform distribution $[-\Delta a, \Delta a]$, and Δa varies as it does for the affine transformation. Each of the four vertices undergoes an independent transformation.

Before the generated shapes can be used in experimentation, they are normalized for translation and scale. Translation is removed by placing the centroid of the shape at zero. The scale is adjusted to one by setting $\sum_i z_i \bar{z}_i = 1$. This forces the denominator of our equation for *Kdist* equal to one.

The shapes were filled with a random noise pattern for several reasons. First, it makes the average gray value in a given patch equal to the background gray value, eliminating total luminance across the shape area as a cue to shape change. This also reduces the afterimages that occur with a uniformly filled shape. Second, the pattern gave a textured appearance to the shape, or a fronto-parallel surface cue, which helps avoid percepts of shapes slanted in depth. Third, the noise reduces the salience of "jaggies" that occur when drawing oblique lines across the display screen.

The stimuli were presented as shown in Figure 4.2. We omitted the use of a fixation marker, instead allowing the subjects to fixate each shape in turn. This allows us to avoid potential side effects of unintended shape distortion due to cortical magnification factors.

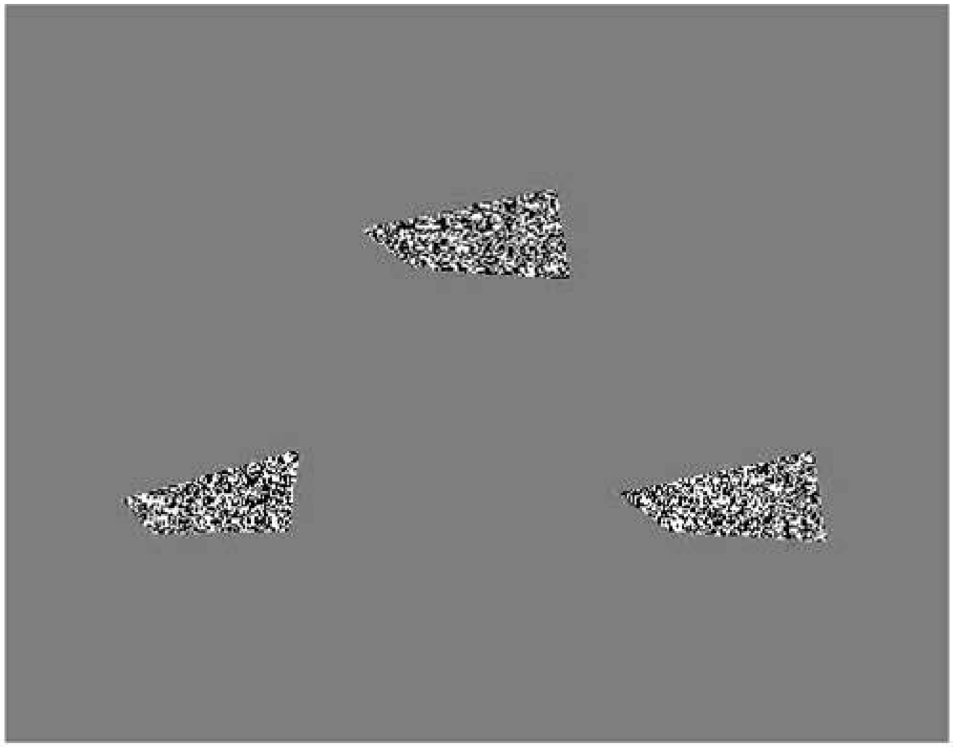

Figure 4.2. The top shape is the reference shape. One of the test shapes (right) is identical to the reference shape and the other (left) has been altered by an affine transformation, exaggerated for illustration purposes.

Procedure

Subjects viewed the stimulus in a dimly lit room to avoid extraneous visual distractions. We follow Goldmeier's paradigm (1972) for judging shape similarity with slight modifications. Subjects are asked to indicate which shape is most dissimilar in a 2AFC task (Figure 4.2). One test shape is always identical to the reference shape, and the other is altered. Subjects were allowed to freely view the stimulus for 4 seconds. The stimulus was followed by a blank white mask for 200 ms. Feedback was given.

Design

A total of 60 reference shapes were created for the experiment, 30 of which were used to measure thresholds for shape discrimination under systematic affine transformations and 30 for the jittered transformation. Only 2 subjects were measured on all 60 shapes, a very time consuming process. The other 2 subjects saw about half of the shapes.

The battery of test shapes (transformed from the reference shape) were compiled in a table and sorted by their distance to the reference shape. To measure thresholds, a dual interleaved staircase was run. The first was controlled by the *Pdist* measure, and the second by the *Kdist* measure. By interleaving the staircases, subjects were unable to predict whether the next trial would be harder or easier based on the feedback they were given on the last trial. The staircase used the "two-down, one-up" method which converges at 70 percent correct performance (Levitt, 1971). We had no information about the underlying psychometric function, so we used a large step size to start, equal to half the standard deviation of all the distance values in the shape table. The starting point for the first trial was a test shape at a distance equal to half the mean of the entire range of the shape table. After 3 and 7 reversals, the step size was cut in half to narrow in on the threshold value, which was taken from data over the next 8 reversals.

Apparatus

The stimuli were presented on a PC running Windows 2000 and the BitmapTools presentation software for Matlab (developed by Payam Saisan, under the supervision of Martin Banks). Subjects viewed the stimulus from 2 meters, a distance at which the screen subtended an angle of 11.31 degrees. The shapes varied in total area, but subtended roughly 4 degrees on average at this viewing distance. The large viewing distance also helped reduce the salience of “jaggies”.

4.3.2. Results and Discussion

The thresholds measured in *Kdist* and *Pdist* for both jittered and affine transformed shapes are shown in Table 4.1 for each subject. At first glance, it looks as if *Kdist* thresholds are more robust to the underlying shape distortion mechanism, making it a better candidate as a perceptual shape metric than *Pdist*.

	Kdist (10^{-2})		Pdist (10^{-3})	
Subject	Jitter	Affine	Jitter	Affine
JMY	3.48 (30)	3.24 (30)	2.08 (30)	3.00 (30)
MDM	3.15 (14)	3.44 (13)	2.14 (14)	1.55 (13)
MYT	2.92 (15)	2.40 (14)	1.66 (15)	1.49 (14)
MJR	2.51 (30)	2.71 (30)	1.48 (30)	2.02 (30)

Table 4.1. Mean thresholds measured for four naïve subjects. The thresholds for measured in *Kdist* appear more consistent for jittered and affine transformed shapes than those measured with *Pdist*. The number in parentheses are the number of quadrilaterals (and independent staircases) tested to obtain that threshold.

We performed a t-test with the null hypothesis that the thresholds are the same for jittered and affine transformations of the shape. The p-values for this analysis can be found in Table 4.2. The values never fall below a significance criterion of $p < 0.05$, so we can not claim that the thresholds are significantly different for either metric.

Subject	Kdist	Pdist
JMY	0.50	0.08
MDM	0.57	0.12
MYT	0.28	0.69
MJR	0.60	0.10

Table 4.2. P-values for a t-test of the null hypothesis that thresholds are the same for random jitter and systematic affine distortions of a shape for each metric.

It is interesting to note that we get the same threshold measured with a given metric, even when the staircase was controlled by the other metric. This lends some confidence that our staircase procedure is converging on a stable threshold value. The 95% confidence intervals for the threshold values can be found in Table 4.3.

	Kdist (10^{-2})		Pdist (10^{-3})	
Subject	Jitter	Affine	Jitter	Affine
JMY	[3.1, 3.8]	[2.6, 3.9]	[1.6, 2.5]	[2.1, 3.9]
MDM	[2.6, 3.7]	[2.5, 4.4]	[1.5, 2.8]	[1.2, 1.9]
MYT	[2.3, 3.5]	[1.6, 3.2]	[1.1, 2.2]	[0.8, 2.2]
MJR	[2.1, 2.9]	[2.1, 3.4]	[1.2, 1.8]	[1.4, 2.6]

Table 4.3. 95% Confidence intervals for the measured threshold values.

The noise or spread we see in our measurement of the thresholds might be attributed to different polygons having different threshold values. If there were true, we would see strong positive correlation across subjects' thresholds for different shapes (Table 4.4). While some correlation exists, it is not consistent across different conditions, so we can conclude that our reference shapes are not systematically different from each other for either metric.

Subject Correlation	Kdist (10^{-2})		Pdist (10^{-3})	
	Jitter	Affine	Jitter	Affine
JMY-MDM	0.67	0.51	0.40	-0.57
JMY-MYT	-0.07	0.66	0.25	0.51
JMY-MJR	-0.07	0.60	0.42	0.64
MDM-MYT	0.07	0.26	0.71	-0.12
MDM-MJR	0.19	0.49	0.55	-0.04
MYT-MJR	0.51	0.69	0.42	0.65

Table 4.4. Correlation between subjects' thresholds for different reference shapes.

Another way to confirm that this is true, is to order the shapes according to their average threshold and see if there is some perceptual shape property that changes with threshold. For example, maybe skinny, vertical polygons have much lower thresholds than fat blocky polygons. Figure 4.3 shows that this doesn't seem to occur.

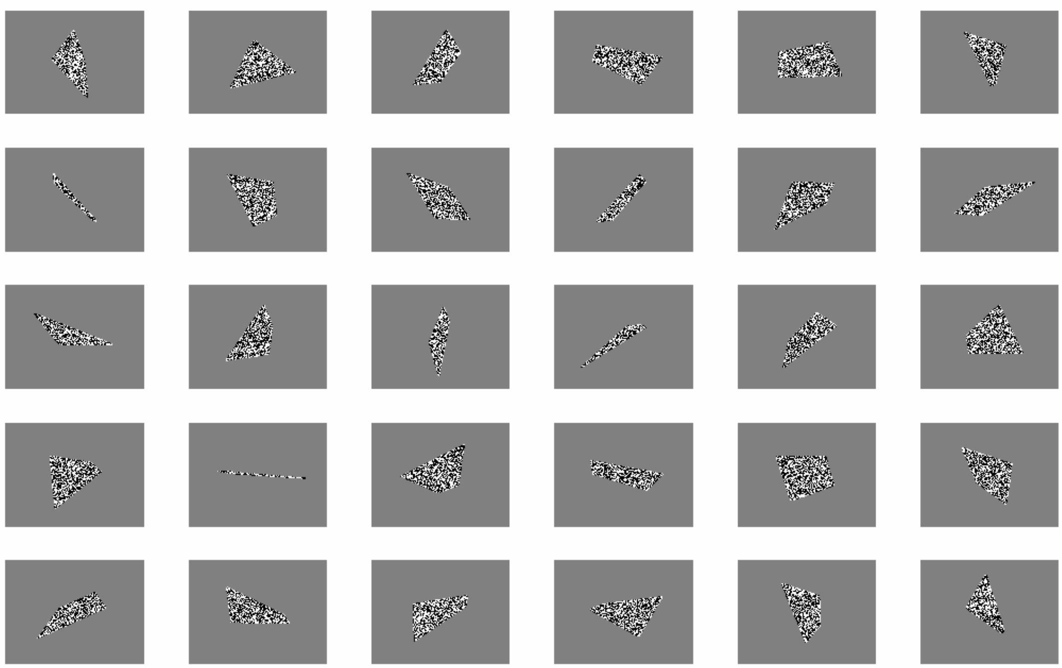

Figure 4.3. Set of reference shapes that were used to measure jnds for affine deformations. The shapes are ordered from left to right, top to bottom, by average *Kdist* threshold measured from the observers. There are no obvious properties that would lead to increasing thresholds in this series.

4.4 Using the metrics

Now that we know we have reliable jnd shape measures, we can see how sensitive subjects' thresholds are to rotation and scale. We know that humans are sensitive to increasing rotation effects, but much less sensitive to changes in scale. We will again measure thresholds to perceived shape change using both *Kdist* and *Pdist.*

4.4.1. Experimental Methods

<u>*Subjects*</u>

One subject from the last study (MJR) and two new naïve subjects participated in these experiments.

<u>*Stimuli*</u>

We create the rotated test shapes by left-multiplying $R(\theta)$ against the reference shape with $\theta = \{0,10,20,30\}$. To create a scale change we apply

$$R(\theta)\begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix}R(-\theta)$$

where $\alpha = \{1,0.8,0.6,0.4\}$ and θ is random. The procedure, design and apparatus are the same as in the previous experiments.

4.4.2. Results

Figure 4.4 shows the average percent change in thresholds for the subjects with decreasing scale. DLW was tested on 2 reference polygons, chosen at random. MJR was tested on 5. Thresholds measured by both metrics show some increase. We can't rule out the possibility that the shapes are getting harder to see. The change in thresholds for rotated shapes is shown in Figure 4.5. *Kdist* thresholds seem to increase greatly while *Pdist* thresholds do not. Given the noise we saw in our threshold measurements, it would be better to measure several subjects on a handful of shapes to better characterize their threshold sensitivity to rotation and scale changes. For this reason, the data in these graphs should only be considered preliminary and no firm conclusions should be drawn at this point. We have provided a rough guide here, however, for future investigations of these metrics.

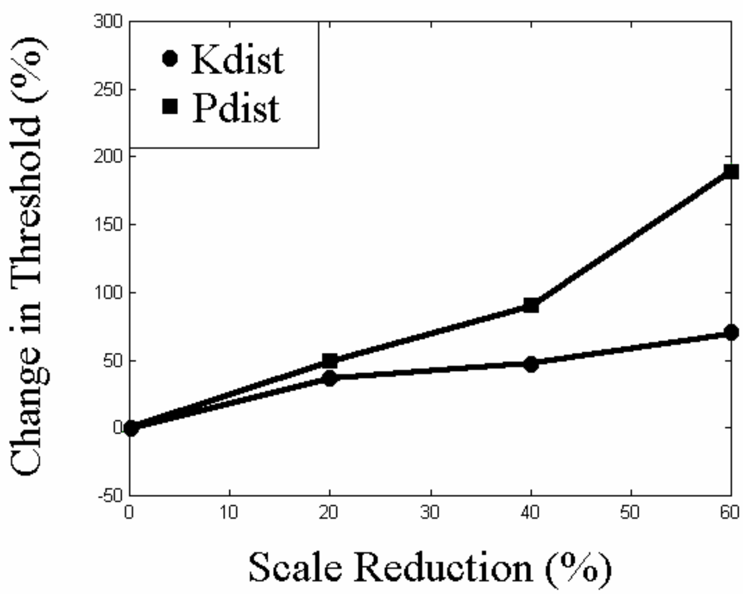

Figure 4.4. Percent change in thresholds to shape perturbations as shapes to be compared are increasingly different in size.

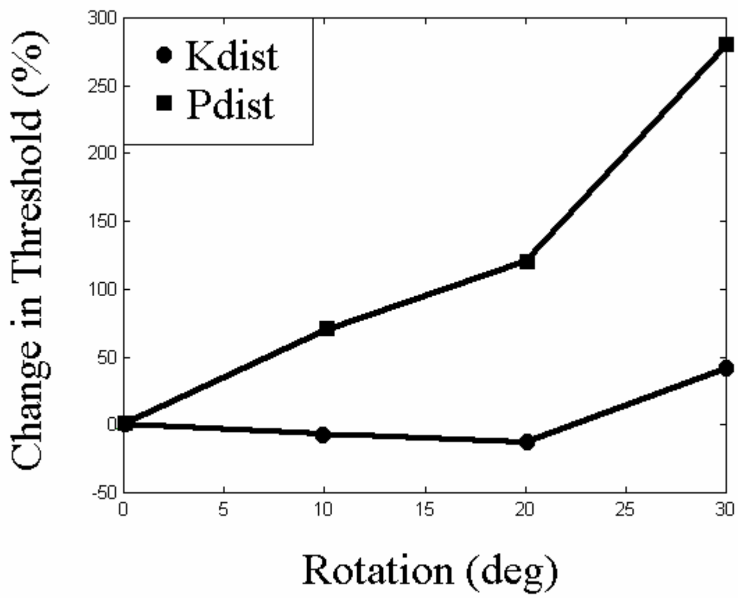

Figure 4.5. *Kdist* thresholds increase as the shape is rotated in the plane. *Pdist* do not appear to increase in these measurements.

4.5 Summary and Future Work

To our knowledge, this is the first attempt at relating quantitative shape metrics directly to human perception. We have shown that both *Pdist* and *Kdist* yield reliable measures of threshold for random and systematic changes to a reference polygon. A next step in this investigation would be to begin moving the test shapes further from the reference shape in the given metric, and have subjects do the same task: determine which shape is more similar. From this, it can be determined how far apart the test shapes must be to get a reliable similarity matching. One possible finding might be that the difference needed between the test shape scales with their distance from the reference shape, or rather, that our perception of shape similarity follows a Weber law. One might want to consider probability summation over orientation of the edges of the polygon as an alternative measure of sensitivity.