A RELATIONSHIP BETWEEN HUMAN SHAPE CATEGORIZATION AND THE STATISTICS OF NATURAL SHAPES.

by

JOHN WILDER

A thesis submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

Dr. Jacob Feldman

and approved by

New Brunswick, New Jersey

January, 2009

ABSTRACT OF THE THESIS

A relationship between human shape categorization and the statistics of natural shapes.

By JOHN WILDER

Thesis Director:

Dr. Jacob Feldman

We studied the classification of shapes into broad natural categories, such as“animal” and “leaf”, into which many shapes can proceed without overt basic-level recognition. Many shape representation models make implicit assumptions about what shape structures often occur in natural shapes, but such assumptions are not generally closely tied to real-world measurements. In order to tune a model of shape classification to the natural environment we collected statistics from a large database of real animal and leaf shapes; first we calculated the MAP skeletons of these shapes, and then computed several different statistical properties of the skeletons. These statistics allow for the creation of an “ecologically-informed” shape classification model that can generalize of many of the specific structures observed in the two classes. To investigate human shape classification subjects were shown shapes created by taking a weighted average of an animal and a leaf shape, resulting in a novel morphed shape. The task was to classify a shape as animal or leaf. Subjects easily performed this task, and their responses were strongly related to the weight used to averaged the shapes. Next, the classifier was used to obtain the likelihoods that a given morphed shape belonged the each class. These likelihood values are predictive of the subjects' responses, suggesting a relationship between human classification of shapes and the shapes' skeletal structures.

Table of Contents

Abstract ii

List of Figures iv

1. Introduction 1

1.1. Natural Image Statistics 3

1.2. Rapid Natural Image Classification 4

2. Experiment 1 6

2.1. Method 6

2.1.1. Subjects 6

2.1.2. Procedure and Stimuli 6

2.2. Results and Discussion 8

3. Experiment 2 10

3.1. Method 10

3.1.1. Subjects 10

3.1.2. Procedure and Stimuli 10

3.2. Results and Discussion 10

4. Statistical Model 12

4.1. Skeleton Estimation 12

4.2. Distribution Estimation 12

4.3. Classifier 15

4.3.1. Feature Selection 16

4.3.2. Model Selection 16

4.4. Fit to the Data 18

4.5. Comparison to Alternative Model 18

5. General Discussion 26

6. Conclusion 27

List of Figures

1.1. Dog and Cat Silhouette 1

2.1. Morphing Shapes 7

2.2. Exp. 1 Subject Responses vs. Morphing Weight 9

3.1. Exp. 2 Subject Responses vs. Morphing Weight 11

4.1. MAP Skeleton 13

4.2. Distributions of Number of Skeletal Branches 14

4.3. Distributions of Unsigned Axis Curvature 15

4.4. Exp. 1 Subjects vs. Naive-Bayes Model 19

4.5. Exp. 2 Subjects vs. Naive-Bayes Model 20

4.6. Naive-Bayes Model vs. Morphing Weight (Exp. 1 Shapes) 21

4.7. Naive-Bayes Model vs. Morphing Weight (Exp. 2 Shapes) 22

4.8. Exp. 1 Subjects vs. Aspect Ratio Model 23

4.9. Exp. 1 Subjects vs. Perimeter^2/Area Model 24

4.10. Exp. 1 Subjects vs. Simple Stats Model 24

1. Introduction

One goal of vision is to help us categorize the objects in the world around us into naturally and functionally relevant classes. One goal of vision science is to find the object feature that will allow for the creation of a model that places an object in its natural category. Many features of an object allow for distinguishing between similar objects or identifying objects. For example, we can distinguish a baseball from a tennis ball because one is white and the other yellow, or we can notice that one is leather and the other is wool. Color and texture cannot, however, are not unambiguous cues and thus cannot reliably classify an object as part of a group (Biederman, 1987). Knowing that an object is yellow does not allow us to classify it as a tennis ball. Seeing something blue does not inform us if we are looking at the sky, a book, or a bird. If our decision rule was to use the feature "blue" to place an object in the category "book" we would find ourselves misclassifying many other blue non-book objects.

Shape, on the other hand, is different. Humans are very good at seeing a shape and using it to identify the object or to determine which category or class of objects it belongs to (Landau, Smith, & Jones, 1988). Shape is also used to identify objects; from only an outline, people are able to identify a rabbit, and even make the fine discrimination between the cat and dog shown in Figure 1.1 (Biederman, 1987).

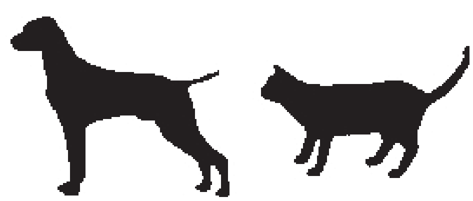

Figure 1.1: This dog and cat can be distinguished from their shape alone.

In the current study, we would like to investigate human shape classification of shape and create a classifier to model the human behavior. If we are to study shape we first need to answer the following question: If shape is used to categorize objects, exactly how is it used? Objects of the same class frequently have similar shapes (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Torralba & Oliva, 2003) so a process that categorizes shapes should place similar objects in one class, and dissimilar objects in different classes. Presumably, each class of objects has a distribution (or several distributions) of features that represents the possible shapes of that class, and this distribution is distinct from that of another class. A key question to answer before finding these distributions is what features need to be represented. For some features this may be fairly intuitive process, but for shape, however, the properties to measure are not obvious. There are many different ways to represent shape, and many different manipulations can be performed on each representation, resulting in a large space of parameters to consider when investigating the classification of shapes into their natural categories.

When choosing a representation one has to consider whether to use representations based on various measures of the shapes outline (contour-based representations) or representations based on larger areas of the shape (region-based representations). Then we decide whether a global (highlighting relationships across the entire shape)approach or a structural approach (highlighting local relationships) will be used. For example, when a structural, contour-based representation is used a set of points from the contour can be selected as the representation. However, it is hard to calculate similarity when using this representation because there are many ways the points could be combined into a single value before the distance between two shapes can be calculated. Alternatively, possible global descriptors include simple shape descriptors, such as area, eccentricity, or principle axis orientation. Using this method multiple descriptors would need to be collected in order to calculate the similarity of two shapes (since two perceptually different shapes could have the same area, or the same axis orientation), but it is difficult to choose which ones are relevant to human shape classification. Categorizing objects could also be based on many other types of representations, including Fourier descriptors (Chellappa & Bagdazian, 1984), boundary moments (M.Sonka, Hlavac, & Boyle, 1993; Gonzalez & Woods, 1992), chain code (Freeman, 1961), grammatical strings (Fu, 1974), shape matrices (Goshtasby, 1985), the convex hull-deficiency (M.Sonka et al., 1993; Gonzalez & Woods, 1992), and shape skeletons (Blum, 1973).

With so many possible representations which should be used in the creation of a classifier? Some of the representations mentioned above, such as boundary moments, do not have an obvious connection to the physical shape and so they will not be used. Others, such as Fourier descriptors, do seem to be perceptually relevant but they have problems when there is partial occlusion or they require a large amount of computation due to high dimensionality (Zhang & Lu, 2004). The medial axis skeleton is a representation that overcomes several of these problems; it is accessible (it can be inexpensively computed from the shape), it is unique for sets of shapes up to an affine transformation, and is sensitive to shape perturbations (Blum, 1973; Marr & Nishihara, 1978). Of the requirements of a good shape representation, as stated in Marr & Nishihara (1978), the medial axis skeleton only lacks stability (small changes in the contour can often drastically change the skeleton). Feldman & Singh (2006) propose a Bayesian method for extracting the relevant skeletal axis, and thus increasing the medial axis' robustness to noise, the call the resulting skeleton the MAP skeleton. There is also a growing set of studies finding psychophysical evidence for the importance of the shape skeleton (Psotka, 1978; Kovacs & Julesz, 1994; Kovacs, Feher, & Julesz, 1998). Additionally, the MAP skeleton highlights the part structure of a shape; the part structure of a shape also has perceptual importance Hoffman & Richards (1984); Hoffman & Singh (1997). For these reasons we chose to use the MAP skeleton as our method of representing the shape when creating a classifier.

After choosing the representation (MAP skeleton) we now need to choose what about the shape skeleton should be measured to create the distributions for different categories. Our approach has its roots in research on natural image statistics.

1.1 Natural Image Statistics

A compelling argument has been made for taking into account the statistical structure of the environment when modeling (Brunswik & Kamiya, 1953; Brunswik, 1956; Gibson, 1966). Including the world's structure in a model provides constraints that can decreases the complexity of that model (Marr, 1982). Using the statistics of the environment, models can be created that do not need to consider all possible worlds, just the ones that are possible for the given environment. This results in models that are more simple and that perform better.

Support for incorporating the structure of the environment in perceptual modeling has been found in studies demonstrating relationships between the structure of perceptual systems and the statistics of the environment (Field, 1987; Geisler, Perry, Super, & Gallogly, 2001). Several studies suggest that these relationships are due to the development of an individual's perceptual system in a certain environment (Hubel & Wiesel, 1970; Hirsch & Spinelli, 1970; Annis & Frost, 1973). The evolution of a species' perceptual systems also seems to be related to the environment (Timney & Muir, 1976; Geisler & Diehl, 2002, 2003). In principle, the structure of the environment could affect both evolution and individual development, making a strong case for taking into account natural statistics when developing a model of perception.

In this paper we will investigate the natural statistics of shape, creating a classifier that models human classification. Many shape perception theories are based on expectations about the way natural shapes are formed (Hoffman & Richards, 1984; Blum, 1973; Biederman, 1987; Leyton, 1989), however an empirical investigation of the statistics of natural shapes has not been carried out. A large database of leaf and animal shapes will be analyzed to obtain statistical summaries of those shapes so that a model of shape classification that is tuned to natural shapes can be created. The goal is a model of shape classification that will be compared to human performance in a shape classification task. Natural categories are organized in such a way that membership in a category varies along a continuum (Rosch, 1973). Human classification of objects into natural categories should then follow a metric that will allow for graded membership. We propose a method to do this using probabilistic classification, revealing the probability that a shape belongs in a class instead of solely an absolute classification.

1.2 Rapid Natural Image Classification

Another motivation for this study comes from studies of humans rapidly classifying natural images. While the current study investigates the classification of natural shapes, previous studies focused on the classification of entire images without segmenting objects from the background. These studies have suggested that the classification of natural scenes happens in a rapid, feed-forward manner (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; VanRullen & Thorpe, 2001; Serre, Oliva, & Poggio, 2007). In these studies the stimulus display time was limited to 20ms yet subjects were still able to identify if a scene contained an animal or not. Thus, classification into animal versus non-animal categories can be successful with only 20 ms exposure to stimuli; VanRullen & Thorpe (2001) suggest that the time required for the visual processing to discriminate between familiar categories takes 150 ms (of course visual processing continues, but this is all that is required by the task). We would like to see if rapid classification can be performed for shapes, as opposed to images of natural scenes.

Here two experiments are reported. In one experiment the participants will examine shapes for as long as they need in order to classify the shape as animal or leaf. In the second experiment the stimulus exposure time will be limited followed by a mask, as in Serre et al. (2007). Finding a difference in the responses of subjects who have limited exposure from those with unlimited exposure would suggest that there are two processes of categorization that can be modeled. One feed-forward model, that is used in cases of brief exposure, and a model of categorization during longer exposure which accounts for feedback from later visual stages.

2. Experiment 1

Subjects were shown novel shapes constructed from natural animal and leaf images. Their task was to decide if each shape was more like an animal or more like a leaf.

2.1 Method

2.1.1 Subjects

Subjects were 28 undergraduates at Rutgers University participating in research projects for course credit. All subjects were naive to the purpose of the experiment.

2.1.2 Procedure and Stimuli

Stimuli were displayed on an iMac computer running Mac OS X 10.4 and Matlab 7.5, using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997).

The stimuli were created by averaging the contours of an animal shape and a leaf shape. The averaging was done by taking each point of the contour with a corresponding contour point of the other shape. First, the principle axes of the shapes were aligned, and then 150 equally spaced points from the contours were sampled and placed in a list. Points in the same position in the list were said to be the corresponding points and a weighted average was computed. From a collection of 250 leaves and 250 animals each possible combination of leaf and animal (62,500 total shapes) was used to create 5 new shapes, each created from a weighted average of the contours, with weights of 30, 40, 50, 60, and 70 percent animal (see Figure 2.1).

Subjects were shown a random sequence of 500 shapes, one shape per trial, and were required to decide whether they thought a shape was more likely an animal or a leaf. A session began with a fixation mark. To start a trial subjects pressed and held down the “A” key and the “L” key. The shape was then displayed and they responded animal by lifting their finger from the “A” key, or leaf by lifting the “L” key. Following a response the screen displayed their response to ensure that they would not forget which key was mapped to each category. Since each shape was neither plant nor animal

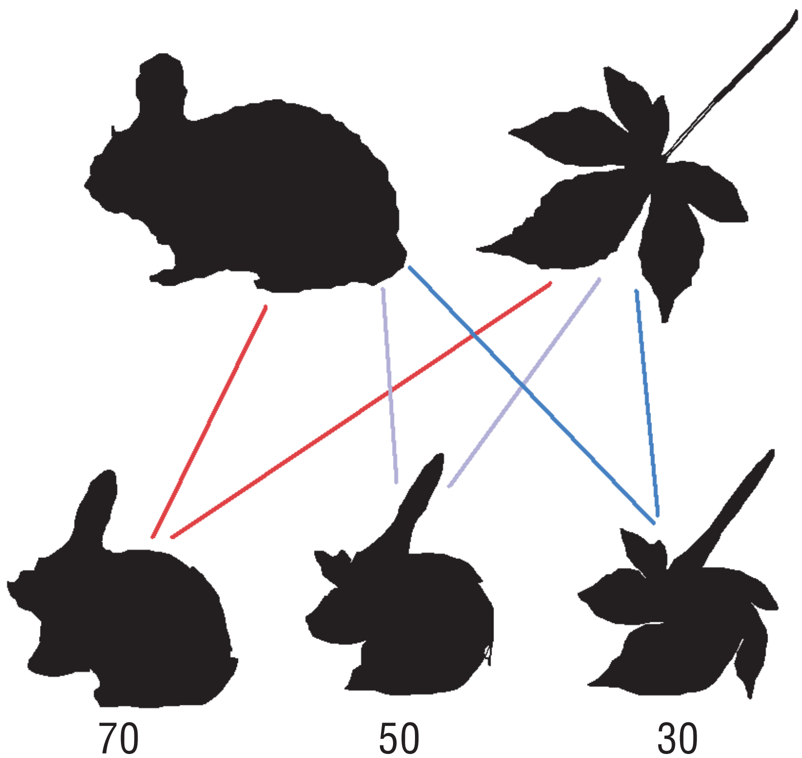

Figure 2.1: One animal shape and one leaf shape averaged into a single combined shape. The contours of the shapes were averaged using the value beneath the shape as as the percentage weighted toward the animal.

there was no correct answer, so feedback informing them if there decision was correct could not be used. The fixation mark was then displayed again so they could proceed to the next trial. The goal was not to see if subjects could perform the task "correctly", but was to discover what shapes subjects perceived to be leaves or animals.

Trials in which the subject responded with both keys were thrown out, and instead of receiving the normal feedback during the trial they were presented a message asking them to only lift their finger from one of the keys. Of the 14000 totals trials 36 were removed.

2.2 Results and Discussion

Subject's responses closely followed the weights used in created the averaged shapes. Figure 2.2 shows the subject's response patterns. There is a strong correlation ($R^2 = 0.64, p \leq 5 \times 10^{-32}$) between the proportion and the weight. When the stimulus was weighted more toward one category, the responses strongly moved in the direction of that category, resulting in a linear fit with a slope larger than one.

Analysis of reaction showed a significant correlation was found, animal responses were correlated with reaction time ($r = 0.7691, p = 0.00002$). However, this high correlation seems to be due to reaction times greater than 10 seconds (≈ 0.57 percent of the trials). If these 80 trials (of the total 13964) are removed from consideration the correlation disappears ($r = 0.2947, p = 0.2352$), suggesting that reaction time is not related to the category of response.

An ANOVA confirmed that the distributions of reactions times were not significantly different across weights ($F_{4,115} = 1.30487 \times 10^{-06}, p = 1$). This is surprising, because one might think that a shape that was weighted more strongly toward a leaf or animal would be more obviously a leaf or an animal, allowing the subject to respond without needing to spend time thinking. This result, however, suggests that it was not more difficult for subjects to categorize shapes that were equally leaf and animal or the reaction times have ceilinged.

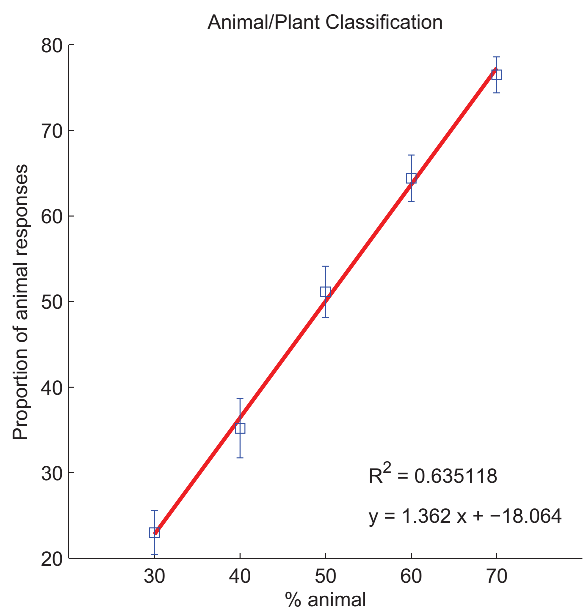

Figure 2.2: A graph showing the proportion of animal responses for each weight used in creating the stimuli. For each subject a proportion was created for each weight. This value was then used to find an average for all subjects, which corresponds to the square in each plot. The error bars are one standard error

3. Experiment 2

3.1 Method

3.1.1 Subjects

As in experiment one subjects were undergraduates from Rutgers who participated in order to receive credit for a psychology course. All 21 subjects were naive to the purpose of the experiment.

3.1.2 Procedure and Stimuli

The stimuli used were the same as in Exp. 1. Each trial began with a fixation mark. Once ready, the subject pressed and held the "A" and "L" keys. The stimulus display length was randomly selected from one of three exposure durations: short (50 ms), medium (100 ms), or long (200 ms). Following the disappearance of the stimulus there was a 12.5 ms blank screen, followed by a mask displayed for 100 ms. The mask was created from shape pieces cut out from the stimulus set. Following the disappearance of the mask the subject would release one of the keys to respond. Their response would then appear on the screen and the next trial would start. A single subject participated in a total of 500 trials.

3.2 Results and Discussion

The three exposure durations (small, medium, and long) did not result in significantly different categorizations ($F_{2,219} = 1.89$ and $p = 0.15$). For this reason the data was collapsed over the three conditions for the remainder of the analysis.

In this experiment we obtained similar results to Exp. 1. The subject's responses closely followed the stimulus weight (Figure 3.1). We see a similar correlation ($R^2 = 0.63$). The main difference to note is the decrease in the slope of the regression line. This reflects that the subjects were not as sensitive to the weight used to create the stimulus (the range of the y-axis in Figure 2.2 is about 20-80, while the range in Figure 3.1 is about 20-65), possibly a sort of speed/accuracy trade-off because processing time

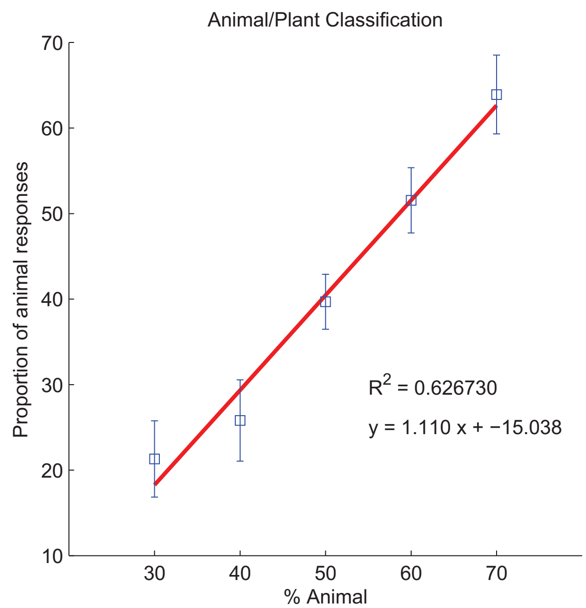

Figure 3.1: The proportion of animal responses for each weight that was used in creating the stimuli. For each subject a proportion was created for each weight. This value was then used to find an average for all subjects, which corresponds to the square in each plot. The error bars are one standard error.

being limited. Also, in Exp. 1 a shape with equal animal/leaf weight was classified as a animal half of the time, but in Exp. 2 subjects only classified these shapes as an animal 40 percent of the time. Overall, with limited stimulus exposure we find that classifications are still strongly related to the manner in which the stimulus was created, however stimuli that were created with an animal bias were less frequently classified as an animal than in Exp. 1.

4. Statistical Model

Many studies of natural image statistics have measured a single property of an image. For example, Maloney (1986) measured the reflectance of objects, Field (1987) measured spatial frequency, and Switkes et al. (1978); Coppola et al. (1998) measured orientation. Shape does not as easily translate into a single measurable parameter as in the previous studies. For reasons described earlier (physical relationship to shape, sensitivity and stability, and psychophysical findings) we used a the MAP skeleton representation of the shapes (Feldman & Singh, 2006) and measured several characteristics of the skeletons.

4.1 Skeleton Estimation

The analysis of the shapes is based on a skeletal representation. The skeleton that maximizes the posterior, $p(SKELETON)p(SHAPE|SKELETON)$, is chosen as the skeleton for a given shape, as described in Feldman & Singh (2006). This results in a skeleton with roughly one skeletal axis per shape part (see figure 4.1).

4.2 Distribution Estimation

After finding a shape's MAP skeleton several characteristics were measured. All of the measured characteristics from animals were used to create an "animal" distribution for that characteristic. The same was done for leaves. After computing the skeleton for each animal and leaf shape the densities, $p(\text{feature} = x|\text{animal})$ and $p(\text{feature} = x|\text{leaf})$, were estimated using histograms.

The different skeletal characteristics measured were: (1) the number of skeletal branches, (2) the depth of the skeleton (using the MAP skeleton's tree structure), (3) the average depth of the branches in the skeleton, (4) the average angle a new branch stems from its parent, (5) the normalized distance along the parent that a child stems, (6) the average branch length relative to the root branch, (7) the average curvature of the skeletal branches (measured in absolute value), and (8) the average branch curvature (taking into account the sign).

We chose to measure the number of skeletal branches in the hope of capturing

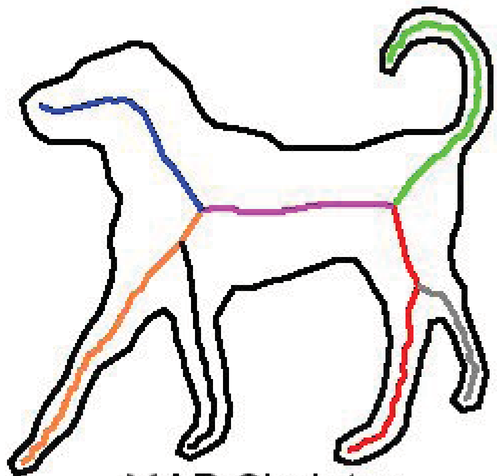

Figure 4.1: A dog shape with its MAP skeleton. Each axis is a different color, and there is roughly one axis per part of the dog

some of the part structure of the shape. The depth of the skeleton should also reveal something about the shape's part structure. The average depth of a skeletal branch is slightly different than the total depth of the skeleton. All three of the statistics mentioned so far together can reveal if a shape has very few branches at each level but a high total depth, or if it has a low total depth but with many branches at each level. We chose to look at the angle a child sprouts from its parent because we noted that animals are able to articulate at joints and leaves do not; this statistics may capture this idea if it is revealed that there is a higher variance in this distribution for animals than plants. The distance along the parent axis that a child branch sprouts may discover if plants and animals generally have their parts in different locations. Finding the average branch length, normalized so the root branch is length 1, will look at if plant branches are shorter or longer than the branches that result in animal appendages. Finally, the last two statistics are the average branch curvatures. This was broken into two different statistics so that a skinny circular shape can be said to be similar and also different than a skinny wavy shape. We can think of two shapes: a snake that is attempting to bite its own tail, and a long sinusoidal branch. They would both have high values when using the absolute value of the angle of curvature, but only the snake would have a high value when looking at the signed value of curvature. The branch would have almost no curvature when taking into account the sign, revealing the overall straightness of the shape.

We can see from Figure 4.2 that there is a large difference between the number of branches on an animal's shape skeleton and a leaf's skeleton. The leaves in our database tend to have a single branch, and the animals more frequently have four or five skeletal branches. This is intuitively correct since we know that animals tend to have four appendages and many leaves are single part shapes, and one of the reasons we chose the skeletal representation of Feldman & Singh (2006) was because this shape skeleton yields roughly one branch per shape part. We also see that the no animal has more than 20 skeletal branches, but there are leaves that have as many as 25. This is due to leaves with many parts, such as a fern leaf.

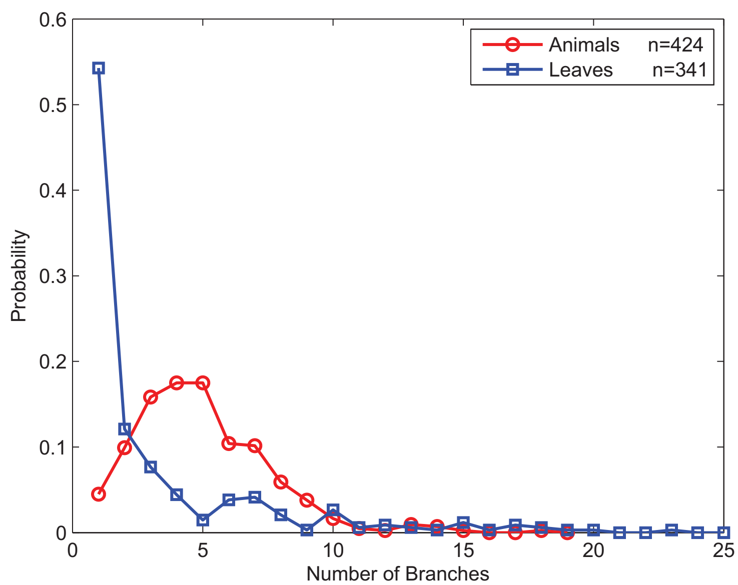

Figure 4.2: The distribution of number of skeletal branches for Leaves (blue) and Animals (red). The x-axis show the number of branches, and the y-axis shows the probability.

Figure 4.3 reveals a smaller difference in the distributions of unsigned axis curvatures than was found in the distributions for the number of skeletal branches, with leaves tending to be slightly more straight than animals.

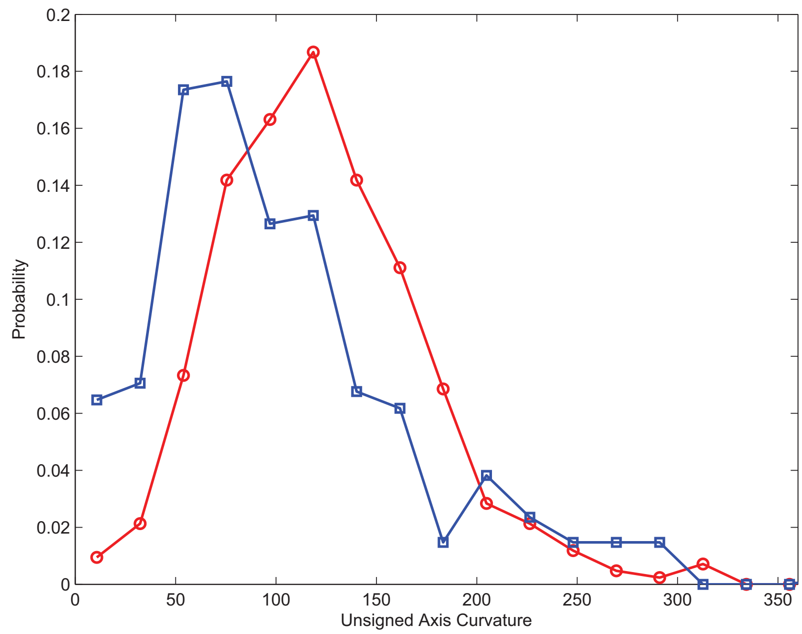

Figure 4.3: The distribution of axis curvatures of shape skeletons for Leaves (blue) and Animals (red). The x-axis show the curvature (in degrees), and the y-axis shows the probability.

4.3 Classifier

A Naive Bayes classifier was created using the distributions measuring the skeletal characteristics as the features used as input. The classifier works as follows:

1. Calculate the skeleton of a new shape x;
2. The skeleton of x will have features $x_1, x_2, \ldots, x_8$; one feature from each of the eight distributions;
3. Calculate the "weight of evidence" according to

$$y = \sum_{i=1}^{8} log \frac{p(x_i|animal)}{p(x_i|leaf)}$$

4. If y is larger than the set criterion the shape should be classified as an animal

4.3.1 Feature Selection

When creating the classifier only features that enhance the ability of the classifier to distinguish between the two classes should be used. For this reason for each feature was only used if a statistical test found a significant difference between the animal and the leaf distribution. Since many of the distributions did not appear to be normally distributed the Wilcoxon rank sum test (Mann-Whitney U) was used (as opposed to the t-test). Three of the features (branch length, distance along parent branch of a child, and the angle a child branches from its parent) failed to be significant using $\alpha = 0.01$, and were not used while creating the classifier.

4.3.2 Model Selection

The classifier was created and tested on the shapes in the training set. We ran the classifier on the training set using each possible combination of distributions in order to find the feature set that is important (we found eight distributions, after feature selection 5 remaining, and there are 31 ways of combining those distributions for use in the classifier). To test which combination of the distributions is best to use as our model we used the Akaike Information Criterion (AIC), which finds the best model while penalizing complexity in order to prevent over-fitting. We computed the AIC according to the following:

$$AIC = 2k - \ln(L)$$

where k is the number of distributions used and

$$L = \prod_{i=1}^{a} L_{animal_i} \prod_{j=1}^{l} L_{leaf_j}$$

and a is the number of animal images, l is the number of leaf images, and for any given animal shape

$$L_{animal} = \prod_{i=1}^{k} \frac{p(x_i|animal)p(animal)}{p(x_i|animal)p(animal) + p(x_i|leaf)p(leaf)}$$

and for a given leaf shape

$$L_{leaf} = \prod_{i=1}^{k} \frac{p(x_i|leaf)p(leaf)}{p(x_i|leaf)p(leaf) + p(x_i|animal)p(animal)}$$

Models with lower AIC values are the preferred models.

We found that the model only containing the distribution for the number of skeletal branches (see the distributions in Figure 4.2) was the preferred model. This model classifies the training set correctly about 79 percent of the time. However, the model that takes into account the number of skeletal branches, the angle a child branch sprouts from its parent, and the unsigned axis curvature, correctly classifies the shapes from the training set about 83 percent of the time. Looking only at the percent correct we may be led to believe that this second model should be chosen, but due to its complexity AIC reveals that the first model should be used. The nine models with the lowest AIC values had a large separated from the other models, but only small difference between them. Since the AIC only shows very small differences between the top nine models we pick the best performing model without fear of over-fitting. Our final model then becomes a model using three features, which are the number of skeletal branches, the unsigned axis curvature, and the angle a child branch grows out from the parent branch.

The procedure was repeated using BIC instead of AIC. BIC gives a larger penalty for increasing the number of free parameters, so it should be no surprise that the computation of BIC suggests that the preferred classifier is the one which uses only the distributions of the number of skeletal branches. However, we see, just as with the AIC, the top nine models have very similar BIC values, so any of the model can be chosen without fear of over-fitting. Finding agreement these two methods of model selection we are confident that we are justified in choosing the three feature classifier described at the end of the previous paragraph.

4.4 Fit to the Data

The morphed shapes used in the experiments were used as input for the classifier. The likelihood/probability values returned by the classifier we used to compare the model to the data collected from the human subjects in Exps. 1 and 2. The likelihood values were binned so that we could calculate the proportion of times the subjects classified the shapes with similar likelihood values as an animal as opposed to a leaf. These proportions were plotted versus the likelihood value the bin represents. A regression line was then fit to the results. This was done for the data collected in Exp. 1 (Figure 4.4) and 2 (Figure 4.5).

For the data in Exp. 1 we find an $R^2 = 0.42, p = 0.003$. So for the data collected in Exp. 1 the models strength of belief that a given shape is an animal is a reasonable predictor of the subject's response.

We find the same trend when looking at the model versus the subject's data from Exp. 2, finding that our regression is very significant $R^2 = 0.78, p = 0.0000006$.

For further confirmation that the model if performing similarly to the human, the model's prediction was plotted against the weight used in the stimulus creation (Figures 4.6 and 4.7). We see that just as the subjects' responses closely follow the morphing weight, so do the predictions of the model.

4.5 Comparison to Alternative Model

In order to show that the fit to the data is not trivial, and that the skeleton is capturing more than just simple low-level shape properties, we also found compared, as a baseline, three other naive-Bayes classifiers to our classifier. One classifier uses the aspect ratio of the shape as its input feature, one uses the ratio of the square of the perimeter to the area, and the final one uses both. There are several reasons these features were chosen for comparison. They are both easy to compute, if the performance of these classifiers is successful there may be no need to spend time computing the skeleton. Also, a large squared perimeter to area may capture information about the part structure of the shape (as does the MAP skeleton); a circular leaf will have a low ratio, while a

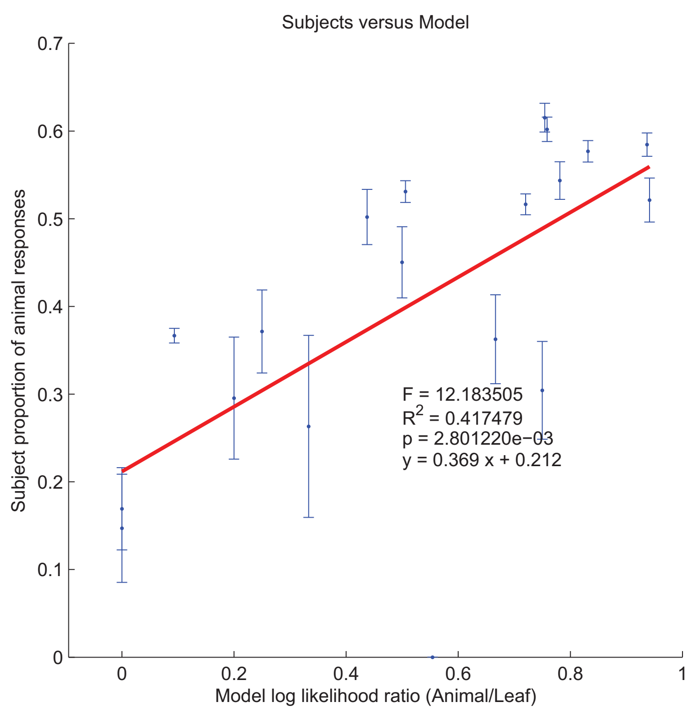

Figure 4.4: The strength of the model's belief that the shape is an animal versus the proportion of times the subjects classified a shape as an animal. This subject data is from Exp. 1 (unlimited exposure duration)

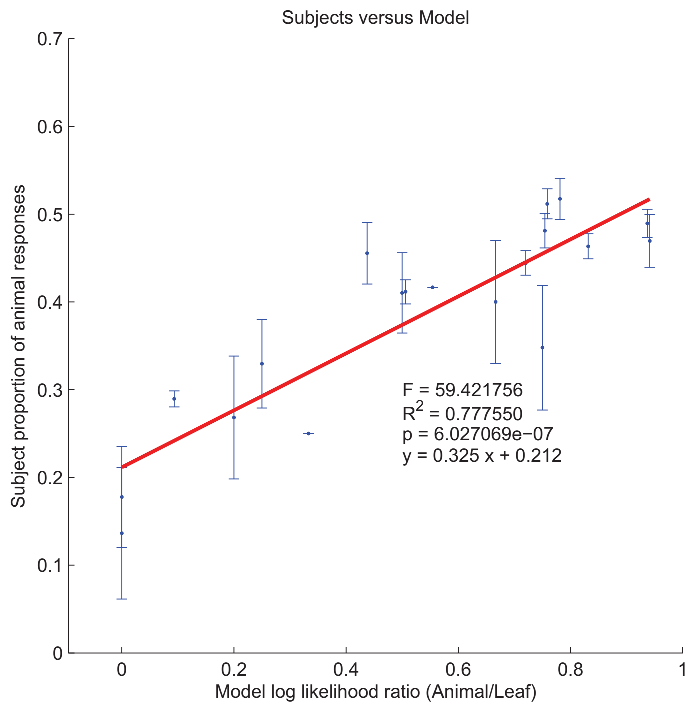

Figure 4.5: The strength of the model's belief that the shape is an animal versus the proportion of times the subjects classified a shape as an animal.The subject data is from Exp. 2 (limited exposure)

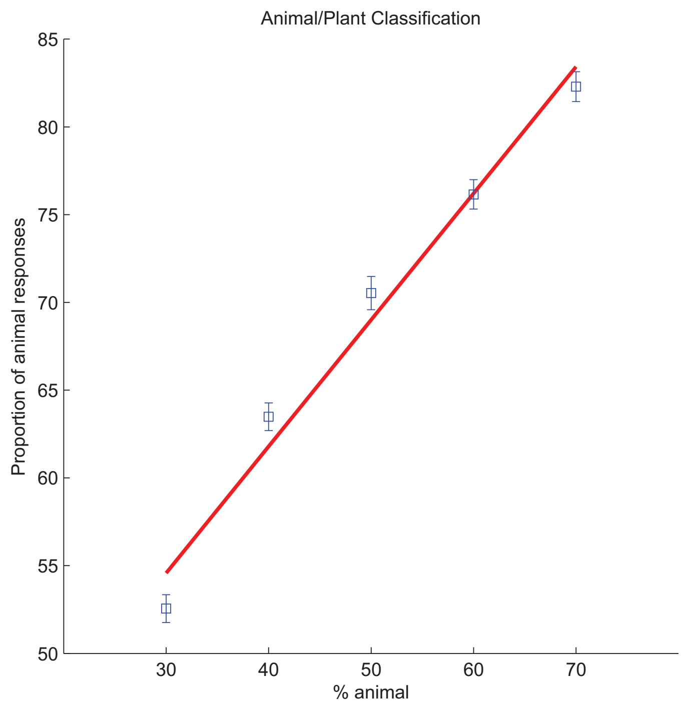

Figure 4.6: The stimulus morphing weight (% animal) versus the proportion of times the classifier would respond "animal". The shapes input to the classifier were the shapes shown to subjects in the first experiment

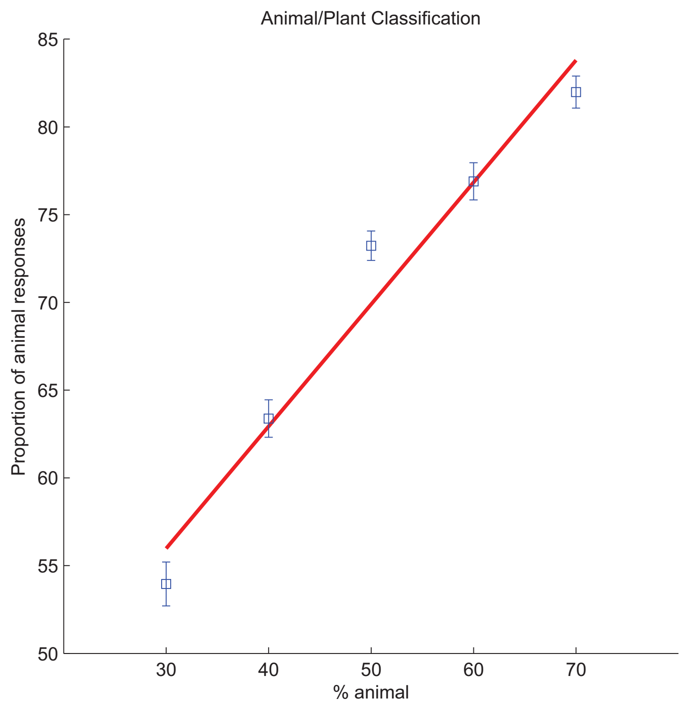

Figure 4.7: The stimulus morphing weight (% animal) versus the proportion of times the classifier would respond "animal". The shapes input to the classifier were the shapes shown to subjects in the second experiment

dog (with area normalized to the leaf's), because of its legs, will have a larger ratio. The aspect ratio may also capture information that the MAP skeleton does, elongated shapes tend to have elongated skeletal axes. A failure of this model (which is expected) will suggest that the skeleton contains important information that is absent in a more simple shape property.

As with the statistics based on the skeletal representation of the shapes, the distributions of shapes for these two features were not normally distributed, so we use the Wilcoxon rank-sum test to test to decide whether or not to use these features in a naive-Bayes classifier. The rank-sum test for both features reveals a significant difference ($p \leq 6.5 \times 10^{-24}$ for perimeter^2/area, and $p \leq 4.6 \times 10^{-7}$ for the aspect ratio), revealing that both features should be useful for inclusion in the classifier.

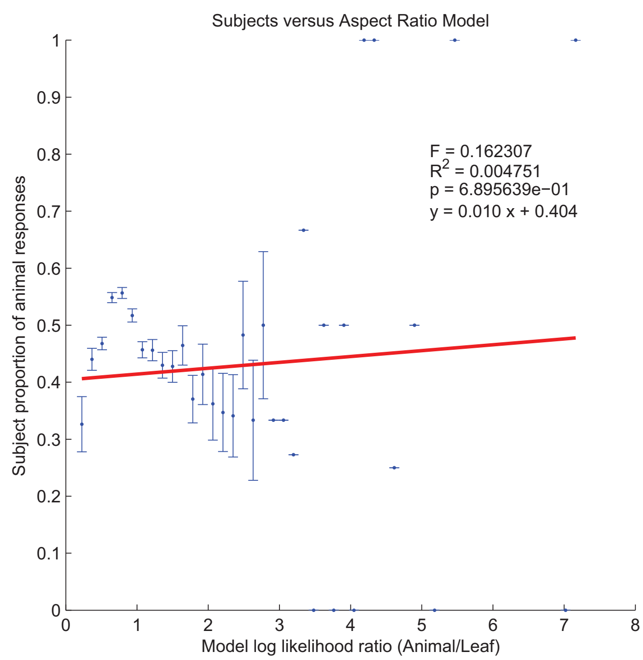

Figure 4.8: The proportion of times the subjects classified a shape as an animal versus the strength of the model's belief that the shape is an animal, the model uses the simple shape statistic Aspect Ratio as its only feature. The subject data is from Exp. 1.

From Figures 4.8, 4.9, and 4.10, we can see these models do not result in the same quality of linear fit as the model based on shape skeleton statistics. These figures were created in the same manner as Figures 4.4 and 4.5; similar likelihood ratios are binned, and for each bin the proportion of times the subjects responded animal was

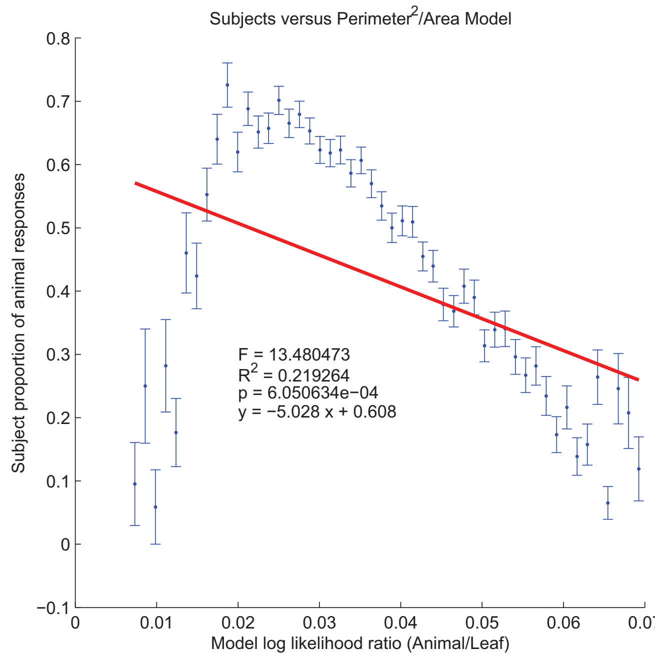

Figure 4.9: The strength of the model's belief that the shape is an animal versus the proportion of times the subjects classified a shape as an animal, the model is using the simple shape statistic Perimeter^2/Area as its only feature.The subject data is from Exp. 1.

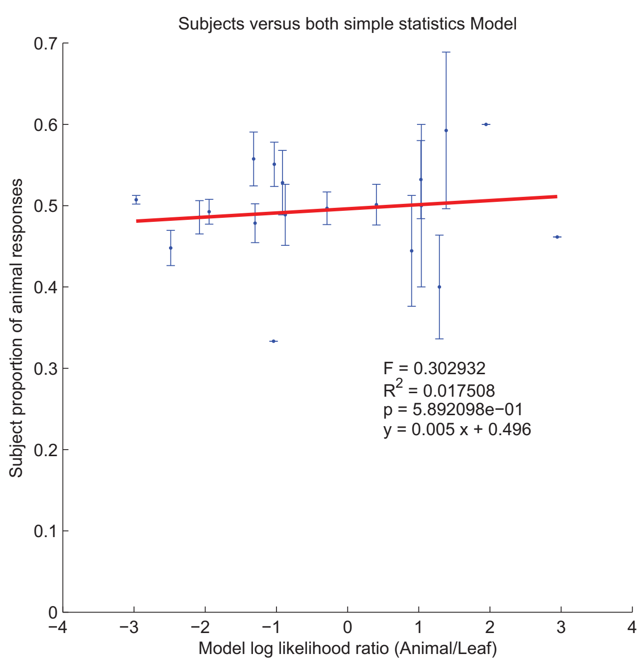

Figure 4.10: The strength of the model's belief that the shape is an animal versus the proportion of times the subjects classified a shape as an animal, the model is using the aspect ratio and the perimeter^2 /area as features. The subject data is from Exp. 1.

calculated. The error bars are one standard error; the points with no error bars too few data points in the bin for a meaningful average and standard error to be calculated. The correlation for the model using the shape's aspect ratio fails to be significant (Fig. 4.8). The correlation for the model using the perimeter squared over area ratio does result in significance, however from Fig. 4.9 it appears that the fit should not be linear, and the correlation was negative. As the likelihood ratio returned by the model increases the model is telling us it more strongly believes that the shape is an animal, this negative regression would suggest that as a shape's fit with the animal distribution increases human subjects are more likely to say it is not an animal. Clearly this is not being done by humans, since they can correctly classify shapes as animals and leaves. Since the shape of the data appears non-linear we could try to perform a non-linear regression, however it does not make sense to do so here, because this would suggest that if the model strongly believes that the shape is an animal or if it strongly believes it is a leaf a human subject would respond "leaf" and if the likelihood ratio does not favor either class more a human would respond "animal". This is also not what humans are doing, since when a human is shown a shape that is obviously an animal (and so its shape statistics suggest it is an animal) the human would not say it is a leaf. Putting both simple statistics together (Fig. 4.10) reveals only a weak correlation that fails to be significant. This suggests that these three simple models do not capture the shape properties used by the human subjects.

5. General Discussion

Initially we proposed Exp. 2 in order to find if a different strategy was used in order to classify the shapes. If there was a different strategy, maybe the same model would not fit the data from both experiments equally well. What we found was only a slight bias for responding leaf in Exp. 2. Our classifier performed well on the data from both Exps. 1 and 2 but we do see that the significance of the regression is weaker when using the data from the first experiment. One obvious explanation for this is that when given unlimited exposure to the stimulus there is no processing limit.

Another possible explanation is if we picked the classifier given the lowest AIC value, classification would solely be based upon the number of skeletal branches of a shape. If classification of a shape as leaf or animal is really (by our subjects) being performed by counting the parts, the limited exposure time may have limited how many parts could be counted, resulting in the slight bias toward leaf classifications that we saw in Exp. 2.

Regardless of the differences in the data from Exps. 1 and 2, what we did see what that just as in rapid natural image classification studies subjects were able to rapidly classify shapes into two natural categories. Also, as in other studies of natural image statistics, we found that the statistical properties of a shape (specifically the statistical properties of the shape's skeleton) can be used to predict or describe how human subjects classify shapes that are presented to them.

The current study shows that a classifier based on the skeleton is consistent with human classifications, adding to the body of work (Psotka, 1978; Kovacs & Julesz, 1994; Kovacs et al., 1998) that suggests the shape skeleton has a roll in the human visual system. More work should be conducted using different classes of shapes and also comparing other alternative models to this one. There are many other classification models that can be created using statistics of the MAP skeleton as features, and future work should investigate these models in order to find the classifier that best models human shape classification.

6. Conclusion

We have shown that humans are able to classify novel shapes into existing shape categories. The poor performance of the models built from simple shape statistics suggests that these shape features are not the most important features that humans use in shape classification, some of the information missing from these simple statistics is highlighted by the MAP skeleton. Models based on the statistics of the shape skeleton are able to better capture the properties of shape that are important for classification, suggesting a roll for the shape skeleton in the human shape classification system.

Bibliography

Annis, & Frost (1973). Human visual ecology and orienation anisotropies in acuity. *Science*, *182*(4113), 729–731.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147.

Blum, H. (1973). Biological shape and visual science (part 1). *Journal of Theoretical Biology*, *38*, 205–287.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkely, CA: University of California Press.

Brunswik, E., & Kamiya (1953). Ecological cue-validity of 'proximity' and of other gestalt factors. *The American Journal of Psychology*, *66*(1), 20–32.

Chellappa, R., & Bagdazian, R. (1984). Fourier coding of image boundaries. *IEEE Transactions on Pattern Analanysis and Machine Intelligence*, *6*(1), 102–105.

Coppola, D., Purves, H., McCoy, A., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, *95*, 4002–4006.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*(2), 171–180.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014–18019.

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America*, *4*, 2379–2394.

Freeman, H. (1961). On the encoding of arbitrary geometric configurations. *IRE Transactions on Electronic Computers*, *EC-10*, 260–268.

Fu, K. (1974). *Syntactic methods in pattern recognition.* New York, NY: Academic Press.

Geisler, W., & Diehl, R. (2002). Bayesian natural selection and the evolution of perceptual systems. *Phil. Trans. R. Soc. Lond.*, *B*(357), 419 – 448.

Geisler, W., & Diehl, R. (2003). A bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, *27*, 379 – 402.

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

Gibson, J. (1966). The senses considered as perceptual systems.

Gonzalez, R., & Woods, R. (1992). *Digital Image Processing*. Reading, MA: Addison-Wesley.

Goshtasby, A. (1985). Description and discrimination of planar shapes using shape matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *7*, 738–743.

Hirsch, H., & Spinelli, D. (1970). Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science*, *168*(3933), 869–871.

Hoffman, D., & Richards, W. (1984). Parts of recognition. *Cognition*, *18*(1-3), 65–96.

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

Hubel, D., & Wiesel, T. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J. Physiol.*, *206*(419-436).

Kovacs, I., Feher, A., & Julesz, B. (1998). Medial-point description of shape, a representation for action coding and its psychophysical correlates. *Vision Research*, *38*, 2323–2333.

Kovacs, I., & Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, *370*, 644–646.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

Leyton, M. (1989). Inferring causal history from shape. *Cognitive Science*, *13*, 357–387.

Maloney, L. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America*, *3*, 1673–1683.

Marr, D. (1982). *Vision*. San Fransisco, CA: WH Freeman.

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

M.Sonka, Hlavac, V., & Boyle, R. (1993). *Image Porcessing, Analysis, and Machine Vision*. London, UK: Chapman and Hall.

Pelli, D. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Psotka, J. (1978). Perceptual processes that may create stick figures and balance. *Journal of Experimental Psychology*, *4*(1), 101–111.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*, 328–350.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *6*, 382–439.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104*(15), 6424–6429.

Switkes, E., Mayer, M., & Sloan, J. (1978). Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis. *Vision Research*, *18*, 1393–1399.

Timney, & Muir (1976). Orientation anisotropy: Incidence and magnitude in caucasian and chinese subjects. *Science*, *193*(4254), 699–701.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*, 391–412.

VanRullen, R., & Thorpe, S. (2001). Is it a bird? is it a plane? ultra-rapod visual categorisation of natural and artifactual objects. *Perception*, *30*, 655–668.

Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, *37*, 1–19.