

Multiresolution Shape Recognition

Eric Chicken and John R. Dixon

Abstract

Shape contours are categorized by analyzing their behavior in the multiscale environment provided by wavelet transforms. Starting with a tangential vector representation of a two-dimensional contour, relevant subspaces of L_2 relating to the contour are identified and used to separate and classify shapes from a well-known test set. The method proposed displays an overall accuracy of 94% for correct shape identification, and attains 100% accuracy for most categories of shapes from the Kimia database.

1 Introduction

One of the main goals of vision analysis is to recognize objects of interest within an image. It is motivated by several applications: face recognition, MRI analysis, and human gait recognition, for example. Analysis and recognition of shape contours is an important part of the overall problem of object recognition in these examples.

Shape contour recognition has been studied extensively. In the past, a common method for examining two-dimensional contours has been a landmark approach (Kendall et al. 1999). In this method, shape boundaries are represented by a finite sampling of the continuous shape contour. It is a difficult process to automate the selection of these landmarks, and often it is done manually, by “expert input”. Additionally, the accuracy of such methods is dependent on the landmarks chosen for analysis.

Le and Kendall (1993), Younes (1999), Grenander (1993), and others approach this problem in terms of shape spaces and manifolds. Here, the shapes retain their continuous properties for analysis (although the contours must by necessity be discretized at some point to be analyzed via computers). These methods were improved on (in terms of computational complexity and a sound statistical framework) by the differential geometry method of Klassen et al. (2004).

We propose here a simple, statistically-motivated method for recognizing shape contours. We use a tangential representation of a closed planar contour and categorize these shape contours using a maximization-penalization technique with respect to the statistical correlation of the shape vectors. Rather than analyzing the entire curve vector, however, we examine instead the projections of the curve into subspaces associated with the multiresolution analysis of the wavelet transform. Only select subspaces are used to classify the shapes, giving us a dimension reduction of our large data set.

By using different scale spaces of the wavelet multiresolution analysis, we take advantage of the ability of wavelets to decompose a signal into decreasingly smooth

approximations. For shape analysis, we expect some of these levels to be similar for different classes of shapes, but, more importantly, we expect that shapes from different classes behave differently in at least one of these approximations. We build our methodology around this assumption, and show that it works very well for distinguishing differing shapes from one another.

Our method does not depend on landmarks, so there are none of the issues present with landmark choice as outlined above. Additionally, the statistical correlation of two shape contours has an intuitive appeal. By using correlation of the contours (actually, the projections of the contours) we tend to identify two shapes that have a strong relation from a particular statistical standpoint.

The use of the wavelet transform is particularly useful in this setting. Using the ability of wavelets to “zoom-in” or “zoom-out” on differing scales of shape features, we are not only looking for strong correlations between two shapes, but looking for strong relations between different scales of the the two shapes.

This paper is organized as follows. Section 2 describes the representation of the shapes under consideration. Section 3 provides a brief description of the wavelets and the subspace projections used to categorize the shapes. Section 4 describes the methodology of our recognition scheme, and the final section gives results from applying our method to a well-known data set of shapes, the Kimia database.

2 Vector representation of shapes

The shapes we consider are closed, planar contours. Given such a shape, an initial point p_1 is chosen arbitrarily on the curve. Another $n - 1$ points are chosen equispaced around the curve with respect to curve length, the direction is counterclockwise. To determine the tangent vector associated with each shape, the following method is used. For each point p_i , α_i is the positive angle between the horizontal line passing through p_i and the line connecting p_i and p_{i+1} . For the n th point p_n , we use p_1 as the follow-on point. An illustration of how these points are determined is given in Figure 1. A complete shape and its associated tangent vector is shown in Figure 2. Our vector representation for these curves is this vector α of tangent angles. For a closed planar contour, the tangent vector will have values in $[m, m + 2\pi)$ for some finite m .

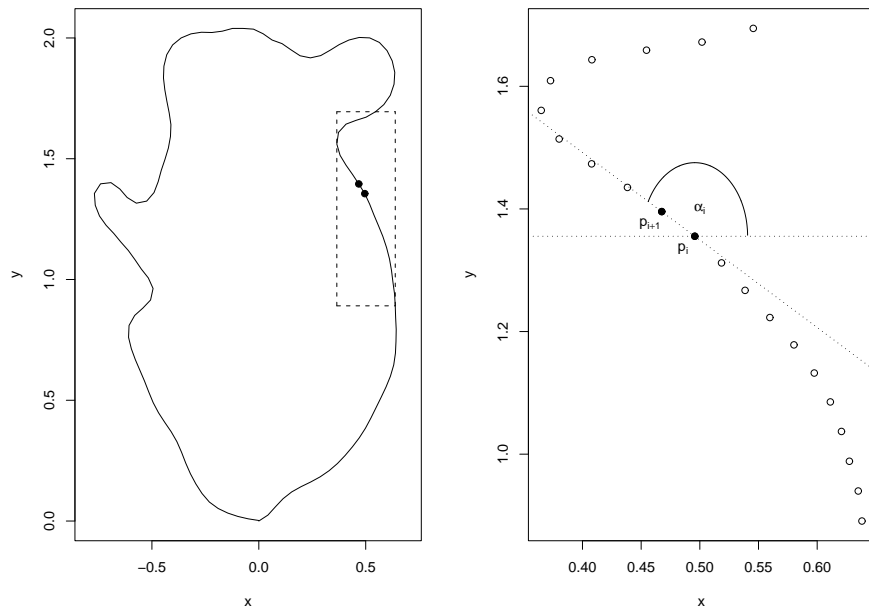


Figure 1: Construction of tangent angle vectors. The left diagram show the entire shape, the right diagram is a closeup of the dashed box shown in the left diagram.

This representation has a few interesting properties. Adding a constant c to the tangent vector is equivalent to finding the tangent vector of the shape rotated by c radians. For simplicity, we will use $c = -m$ for all our shape tangent vectors. Shifting the angle vector (modulo 2π) to the left or right is equivalent to changing the starting point p_1 for the contour. See Figure 3.

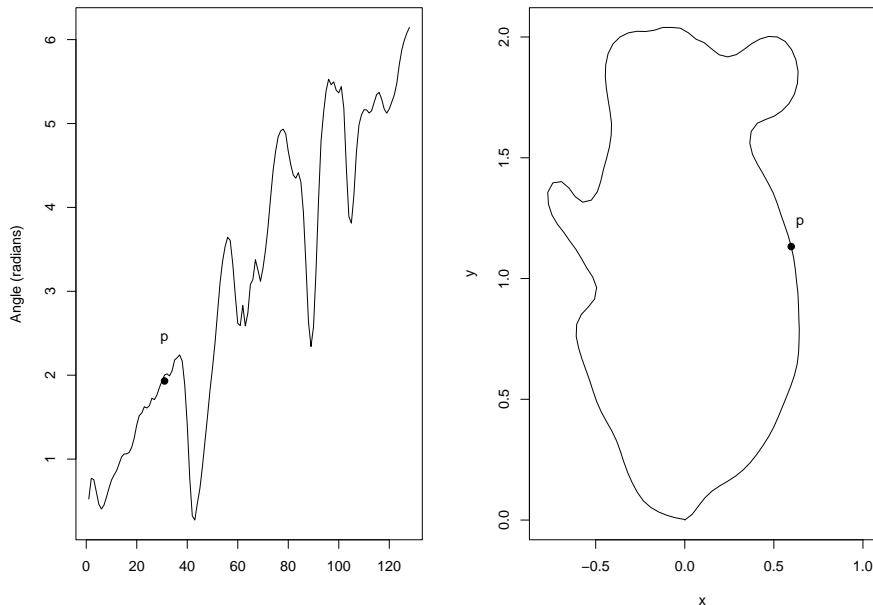


Figure 2: Example of a modified tangent angle vector and its associated planar contour.

3 Wavelets and multiresolution spaces

We examine the shape tangent vectors indirectly by analyzing their projections onto subspaces associated with the multiresolution analysis (MRA) of wavelets.

Wavelets are very efficient at function estimation and compression. They excel at spatial adaptivity, optimality in terms of error rates, and low computational cost. Usually, wavelet analysis is performed through the use of thresholding or denoising the wavelet coefficients, such as the VisuShrink method of Donoho and Johnstone (1994), or the block methods analyzed in Chicken (2003, 2005), Cai (1999) and Chicken and Cai (2005). There, a noisy signal is transformed into empirical wavelet coefficients by the discrete wavelet transform (DWT) of Mallat (1989), these coefficients are denoised by comparison with a specified thresholding rule, and the underlying function is estimated by applying the inverse DWT to these denoised coefficients.

Wavelets have the useful property that they can simultaneously analyze a signal f in both time and frequency. This is accomplished by projecting f into function spaces of varying degrees of smoothness. This is referred to as the multiresolution property of wavelets. In usual wavelet analysis, the different projections are summed together

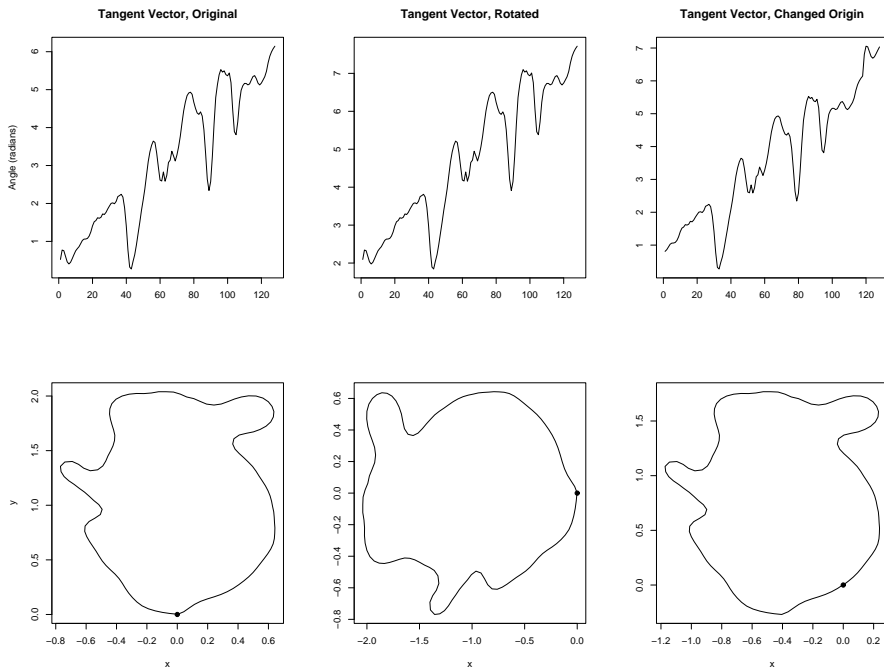


Figure 3: Illustration of properties of tangent angle vectors. The origin is marked in each of the lower diagrams.

after some modifications involving compression or denoising. In this paper, we are more interested in the behavior of the function at these different resolution levels rather than reconstructing the original function.

3.1 Wavelets

Wavelets are an orthogonal series representation of functions in $L_2(R)$. Let ϕ and ψ represent the father and mother wavelet functions, respectively. There are large families of choices for these two functions available for use, see for example Daubechies (1992). Here, we choose ϕ and ψ to be compactly supported and to generate an orthonormal basis. For example, see Figure 4. Let ϕ_{jk} and ψ_{jk} be the translations and dilations of ϕ and ψ :

$$\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k),$$

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k).$$

Then for any fixed integer j_0 ,

$$\{\phi_{j_0k}, \psi_{jk} | j \geq j_0, k \text{ an integer} \}$$

is an orthonormal basis for $L_2(R)$. Let ξ_{j_0k} and θ_{jk} be the usual inner product of a function $f \in L_2(R)$ and the wavelet basis functions:

$$\xi_{jk} = \langle f, \phi_{jk} \rangle,$$

$$\theta_{jk} = \langle f, \psi_{jk} \rangle.$$

Then f can be expressed as an infinite series

$$f(x) = \sum_k \xi_{j_0k} \phi_{j_0k}(x) + \sum_{j=j_0}^{\infty} \sum_k \theta_{jk} \psi_{jk}(x).$$

Since we do not have the entire, continuous tangent function f , but merely have a discrete realization of it, we estimate these wavelet coefficients using the DWT. If f is represented as a vector of dyadic length $n = 2^J$ for some positive integer J , then the DWT will give us a total of n estimated coefficients ξ_{j_0k} and θ_{jk} over the indices $j = j_0, j_0 + 1, \dots, J - 1$ and for all appropriate k . The lowest level possible for j_0 is 0, but higher values may be chosen.

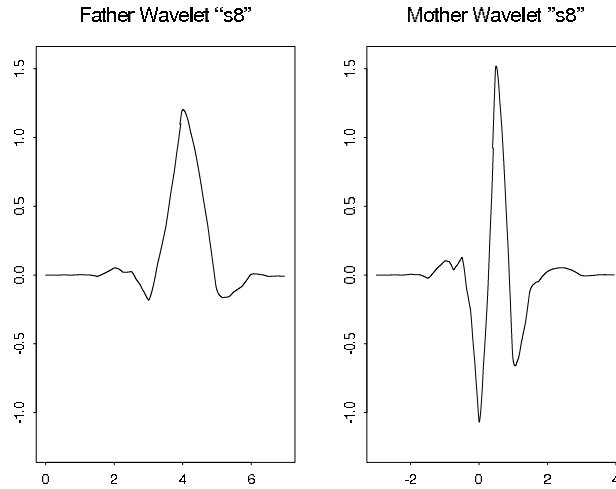


Figure 4: Example of father and mother wavelets functions ϕ and ψ .

The series representation for f above can be viewed as several series: one series involving the father wavelet at level j_0 , and one mother wavelet-based series for each $j \geq j_0$. The first series corresponds to the smooth, or coarse, structure of f . The other series correspond with increasingly detailed, high-frequency parts of f . Changing the index (or resolution level) j allows us to zoom in or out onto the smooth or detailed structure of f . This is referred to as the multiresolution property of wavelets, and allows the wavelets to simultaneously analyze a signal in both time and frequency.

3.2 Multiresolution Analysis

Each of the series at a different resolution level j corresponds to a subspace of $L_2(R)$. Let V_j be the space spanned by the ϕ_{jk} and W_j be the space spanned by the ψ_{jk} . For

wavelets as constructed by Daubechies, V_j is orthogonal to W_l for $l \geq j$. Additionally, there is a ladder structure to these subspaces. If $f \in V_j$, then $f(2\cdot) \in V_{j+1}$. This creates a ladder of spaces:

$$\cdots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \cdots \subset L_2(R)$$

with $\overline{\cup_j V_j} = L_2(R)$ and $\cap_j V_j = \{0\}$. For large j , then, V_j is an approximation of the space $L_2(R)$. Each space V_j is “closer” to $L_2(R)$ in that it contains functions of increasing complexity.

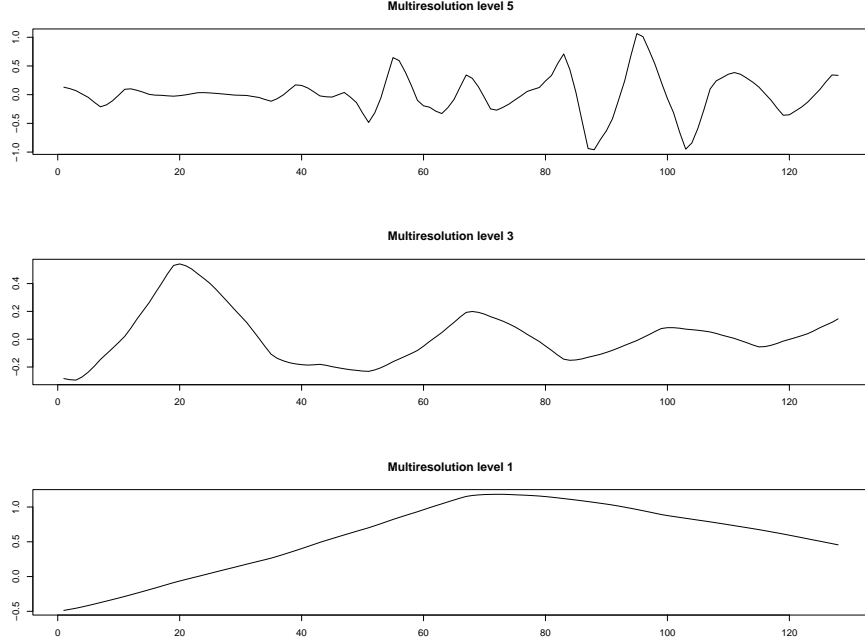


Figure 5: Projections of the tangent function for the rabbit shape onto three resolution spaces.

The mother wavelet ψ is constructed so that not only is V_j orthogonal to W_l for $l \geq j$, but

$$V_{j+1} = V_j \oplus W_j$$

Using this, we choose a large value $j = J$ and use V_J as the approximation of $L_2(R)$. Then

$$V_J = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \cdots \oplus W_{J-1} \quad (1)$$

Each space V_{j_0} and W_l in the above expression represents parts of an $L_2(R)$ function at different resolutions. The wavelet transform projects the function f into each of these spaces. In this paper, we are interested in the behavior of the function at these different resolution levels. In particular, we will look for strong statistical correlations between corresponding resolution levels of different shape tangent functions.

As an example, consider the tangent vector for the rabbit shape seen in Figure 1. Using the father and mother wavelets shown in Figure 4, three of the projections are shown in Figure 5. Since the tangent function is of length 128, then $J = 7$. Using $j_0 = 0$,

the DWT provides resolution levels j_0 through $J - 1$. The top plot is the projection of the tangent vector onto W_5 , the middle panel is the projection onto W_3 and the bottom panel is for W_1 . The original function can be reconstructed by adding all the projections from j_0 to $J - 1$ together.

Note that in this figure, each of the projections becomes more detailed as the resolution level increases. The bottom panel is the smoothest projection, and the projections become increasingly more detailed in the middle and top panels.

4 Methodology

Suppose we have m categories of shapes and wish to take a shape B and determine which of these categories it belongs in. For each category, we will compare shape B to known members of that category. From our set of contours, we randomly select five members from each category to act as a training set. We will use this training set to determine a match between the unknown shape B and a particular category.

Several things are needed to determine a match. First, we need a metric or score for quantifying the resolution level comparisons. Second, we need to know which resolution levels to compare. Lastly, we need to decide when a match occurs. We consider each of these tasks in turn.

4.1 Score and penalty function

We form a match by observing the correlations among the different resolution levels between a known training set shape and an unknown shape. Each MRA level is a vector of length 128 representing the projection of the tangent vector onto one of the subspaces in (1). The correlation we use is the usual Pearson's coefficient of correlation:

$$\rho = E(X - \mu_X)(Y - \mu_Y) / \sqrt{\text{var}(X) \cdot \text{var}(Y)}$$

where X and Y are random variables. ρ measures the degree of linear relation between X and Y . Values near ± 1 imply a linear relation (with positive/negative slope) between the two, and lesser values show a lesser strength relation. We estimate ρ with a random samples x_i and y_i of size n having the same distributions as X and Y with

$$\hat{\rho} = \frac{\sum_{i=1}^n x_i y_i - \left(\frac{1}{n} \sum_{i=1}^n x_i\right) \left(\frac{1}{n} \sum_{i=1}^n y_i\right)}{\sqrt{\left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \frac{1}{n} \sum_{i=1}^n x_i)^2\right) \left(\frac{1}{n-1} \sum_{i=1}^n (y_i - \frac{1}{n} \sum_{i=1}^n y_i)^2\right)}}$$

Instead of X and Y , we use the projection at resolution level j of a shape in the training set and resolution level j of the unknown shape B . We view them as random in the sense that the shapes and their tangent vectors are coming from a population with a mean shape, but each observed shape varies from this mean in some fashion.

Assume that we will be using k resolution levels $j = j_1, j_2, \dots, j_k$ to compare the two shapes. The selection of the k levels is explained in the next subsection. Let $\hat{\rho}_i, i = 1, 2, \dots, k$ be the correlation between the resolution levels j_i for each shape. If two shapes are exactly the same (or perhaps reflections of each other), then $|\hat{\rho}_i|$ should

be 1 assuming that the tangent vectors begin at the same point. Unfortunately, this is not a valid assumption for the shapes due to the fact that the shapes all vary around the ideal case.

If the tangent functions do not have the same starting position on the shape, then each resolution level will be out of phase, also. As shown in Figure 3, rightmost panel, shifting the tangent vector will move the origin. So, we may shift the tangent function to align the starting points. Since we are going to match shapes by using the projections onto subspaces rather than the actual tangent vector, we shift the projections of the resolution levels instead. For each resolution level j_i , we choose the shift s_i that gives the maximum correlation $\hat{\rho}_i^*$ between the two shapes at this particular level. The sum of these maximum correlations over the resolution levels after shifting is our initial measure of how similar the shapes are. For two shapes A (from the target training set) and B , we write this as the score for the match,

$$r(A, B) = \sum_{i=1}^k |\hat{\rho}_i^*| \quad (2)$$

Since correlations are always in $[-1, 1]$, this measure resides in $[0, k]$. For two shapes that are similar, the shifts s_i at each resolution level will be about the same, the correlations $|\hat{\rho}_i^*|$ will be near 1, and $r(A, B)$ will be near k .

However, if the two shapes are different, then shifting resolution level j_i by an amount s_i to maximize the correlation at that level is not measuring how close the two shapes are. Instead, it is merely finding a shift where the two levels appear most similar. It is possible that this value can be large even if the shapes are quite different. However, although this may be case, the shift that maximizes the correlation at one level will not do so at other levels. To take this into account, we introduce a penalty on the correlation score r at (2).

For a particular level j_i and an optimal shift s_i , we look at the correlations of the other resolution levels for that shift. Let $\hat{\rho}_{i,l}^*$, $l \neq i$, be the correlation between resolution levels j_l for the two shapes when these levels are shifted by the amount s_i . If the shapes are not similar, then the shift that maximizes the correlation at level j_i will not maximize it for another level. These other correlations are used to penalize the sum of the maximum correlations at each resolution level. Our penalty is determined by taking the absolute value of the difference between the correlation $\hat{\rho}_i^*$ at level j_i (the level we intentionally maximized) and the highest correlation among the remaining resolution levels $\hat{\rho}_{i,l}^*$, $l \neq i$:

$$p(A, B) = \sum_{i=1}^k p_i(A, B) = \sum_{i=1}^k \left| |\hat{\rho}_i^*| - \max_{l \neq i} |\hat{\rho}_{i,l}^*| \right| \quad (3)$$

Note that $p_i(A, B) \in [0, 1]$ and $p(A, B) \in [0, k]$. Our measure of similarity for two shapes is then the score from (2) combined with the penalty (3),

$$r(A, B) - p(A, B) \in [-k, k]. \quad (4)$$

When performing actual comparisons of a shape B to the shape categories, we obtain the score at (5) for each of the m categories.

We actually have five training shapes from each category, so we can perform five comparisons between B and any category. We find five scores $r(A_j, b)$ for each A_j in the training set for a particular category. We use the maximum of the the five scores as our total score for the comparison of B to that category.

$$\max_{1 \leq j \leq 5} (r(A_j, B) - p(A_j, B)) \quad (5)$$

4.2 Determining resolution levels

The DWT provides us with one coarse resolution level ($j_0 = 0$) and $J = 7$ detail resolution levels ($j = j_0 = 0, 1, \dots, 6 = J - 1$) associated with the subspaces in (1). We have chosen to use the Haar wavelet for this analysis. For the Haar wavelet, $\phi = 1_{[0,1]}$, the indicator function on the unit interval. The mother wavelet is then

$$\psi(x) = \begin{cases} 1, & x \in [0, 1/2) \\ -1 & x \in [1/2, 1) \\ 0, & \text{otherwise} \end{cases}$$

We chose the Haar wavelet in part because of its ease in dealing with boundary conditions (both the DWT and the Haar wavelet are confined to the unit interval). Other wavelets may, of course, be used. With the Haar basis, the coarse level associated with V_0 is a horizontal line for all functions, and the lowest detail resolution level (W_0) is multiple of ψ above. For the projection onto V_0 , the score function is always 1 and therefore contains no information about matching the two shapes. For W_0 , the score function will be either 1 or -1 . Since the tangent vectors are generally increasing (see Figure 6), this is nearly always positive 1. Therefore, these levels are never considered for inclusion in the score. Additionally, the highest resolution level (W_6) mainly models noise, not shape structure. This level is also removed from consideration.

4.3 Declaring a match

To determine a match, we compare a candidate shape B to the known shapes from each category's training set. We decide that the candidate belongs to the category where the highest score-penalty (5) was obtained. This method guarantees that every candidate shape will be placed in a category. If it is known that the candidate shape B must belong to one of the m categories, then this method is reasonable. However, if the situation is such that the candidate shape need not belong to one of the categories, then this is not the best method.

In this latter case, it would be better to declare a match with the category with the highest score, provided that it is larger than some minimum threshold needed for a match. Using this method, a shape that does not belong to any category may not be matched with any category. This cannot happen in the previous setup.

To determine this lower threshold, we return again to the training data set. Using the resolution levels selected before, we determine the minimum score for a match by comparing all shapes within each category's training set. Each set contains five shapes, so there are ten possible comparisons to make within each category, and we obtain ten scores as at (4). Each training set shape in a category is compared to the four remaining

shapes from that training set, and we note the median of these four measurements. We do this for each of the five test shapes, resulting in five medians.

Given these five scores for known matches within the training set, we determine a minimum threshold by making as few assumptions on the distribution of the scores as possible. We therefore assume that for a given category, the scores are uniformly distributed over some interval $[a, b]$. We would like to declare a match whenever the the score is at least a . Using this non-informative assumption of uniformity, the estimate of a is

$$(5x_{(1)} - x_{(5)})/4$$

where $x_{(1)}$ is the minimum of the five scores from the training set for a category, and $x_{(5)}$ is the maximum.

Since this estimator of a depends on the range of the scores, and the range of the scores depends on the number of scores, this estimate is upwardly biased. To account for this, we multiply the lower threshold given above by 0.9. Therefore, our minimum threshold needed to declare a match for a category is the score

$$9(5x_{(1)} - x_{(5)})/40 \tag{6}$$

5 Simulations on test data

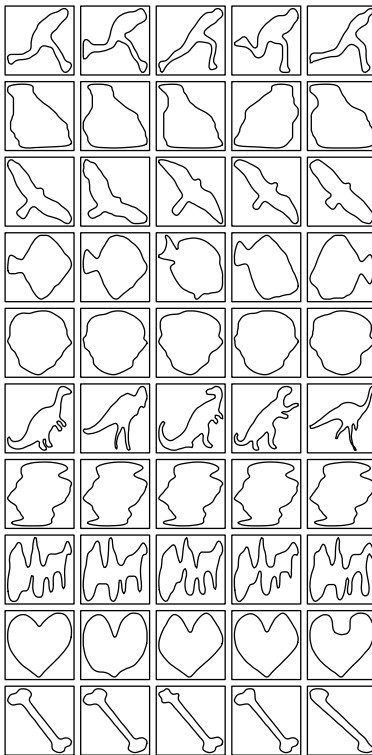


Figure 6: Five training shapes from ten categories of the Kimia database.

We randomly selected ten categories from the Kimia database to test our method on. From each of the ten categories, we randomly selected five shapes to represent the

training set, and ten shapes to test the method. Figures 6 and 7 show the training set and data set, respectively.

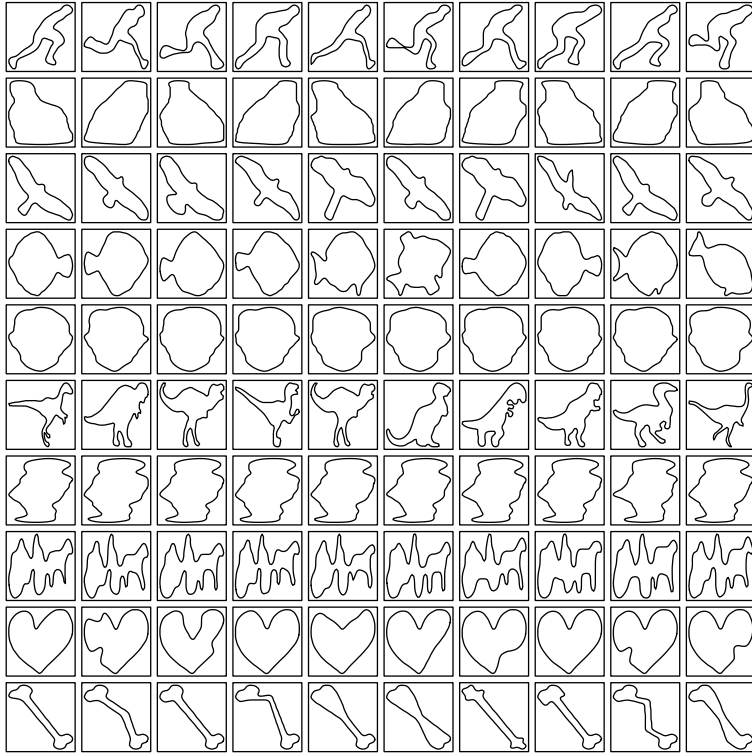


Figure 7: Ten shapes from each of ten categories of the Kimia database.

We ran the method on the test shapes using the methodology described in Section 4. For each of the 100 test shapes B , we computed the scores at (5) for each of the $m = 10$ categories of shapes. Table 1 shows the thresholds found using the method described in Section 4.3. The match is declared by taking the largest of the ten scores at (5) that is above its particular threshold. Table 1 shows the percentage of correct category matching for these 100 shapes and ten categories using this method.

Of the 100 shapes examined, all were placed in some category. There were no shapes that couldn't be classified. Non-classification is a possibility if no candidate shape B attained a score greater than any of the thresholds from Table 1.

Eight of the categories had all ten test shapes matched to the correct category using this method. The remaining two categories, "Flatfish" and "Dinosaur", had two and 4 mismatches, respectively. One of the "Flatfish" shapes was matched incorrectly to the "Carriage" category, and the other to the "Dinosaur" category. Two of the "Dinosaur" test shapes were matched incorrectly to the "Bird" category, and two to the "Bottle" category. In four of these six misspecifications, the correct category was the next highest score. For the other two errors, the correct category was the third highest score.

A possible explanation for the poor performance of the "Dinosaur" category may be seen by observing the shapes in Figures 6 and 7. There it can be seen that the shapes in this category show much more variation than the others. In particular, there are notable differences in body type (thick vs round), leg type (long vs short), and neck

<i>Shape</i>	<i>Threshold</i>	<i>% Match Correct</i>
Step	3.52	100
Bottle	2.54	100
Bird	3.15	100
Flatfish	3.26	80
Face	3.65	100
Dinosaur	2.47	60
Child	3.86	100
Carriage	2.72	100
Heart	3.77	100
Bone	3.44	100

Table 1: Thresholds and match percentages for 10 categories from the Kimia shape database.

type (long vs short). The small training set cannot accurately reflect this diversity. For example, it contains only one of the long neck types. A larger training set would certainly have a higher probability of including more of the variations of this shape category. Also, it may be prudent to break the “Dinosaur” category into to multiple categories based on the differing shape types.

6 Remarks

There are two possible refinements to this method that could be considered. First, the number of resolution levels might be reduced further. We currently use five of the available eight. Examination of the training set could point out which of these five levels are most important for detecting similarities and differences among or between the shape categories. Using fewer than five levels would provide for a greater dimension reduction to the problem.

Second, a larger training set would improve the results. Not only would it give us a better estimate of the lower bound for thresholding derived in Section 4.3, but it likely improve the match percentages for categories that perform poorly under the current algorithm. If there were more shapes in each training set, then the score-penalty function would have a greater range. A greater range would facilitate differentiation among the categories.

We note that our results in terms of shape recognition reported in Section 5 are quite strong. Implementing the suggestions above may not result in enough improvement to be worth considering,

References

- CAI, T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Statist.* **27** 898–924.
- CHICKEN, E. (2003). Block thresholding and wavelet estimation for nonequispaced samples. *J. Statist. Plann. Inference* **116** 113–129.
- CHICKEN, E. (2005). Block-dependent thresholding in wavelet regression. *Journal of Nonparametric Statistics* **17** 467–491.
- CHICKEN, E. and CAI, T. (2005). Block thresholding for density estimation: local and global adaptivity. *Journal of Multivariate Analysis* **95** 76–106.
- DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.
- DONOHO, D. and JOHNSTONE, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81** 425–455.
- GRENANDER, U. (1993). *General pattern theory*. Oxford University Press.
- KENDALL, G., BARDEN, D., CARNE, T. and LE, H. (1999). *Shape and shape theory*. Wiley.
- KLASSEN, E., SRIVASTAVA, A., MIO, W. and JOSHI, S. (2004). Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Pattern Analysis and Machine Intelligence* **26** 372–383.
- LE, H. and KENDALL, D. (1993). The Riemannian structure of Euclidean shape spaces: a novel environment for statistics. *Annals of Statistics* **21** 1225–1271.
- MALLAT, S. (1989). Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.* **315** 69–89.
- YOUNES, L. (1999). Optimal matching between shapes via elastic deformations. *Journal of Image and Vision Computing* **17** 381–389.