

Shape L'Âne Rouge: Sliding Wavelets for Indexing and Retrieval

Adrian Peter¹, Anand Rangarajan² and Jeffrey Ho²

¹Dept. of ECE, ²Dept. of CISE, University of Florida, Gainesville, FL

Abstract

Shape representation and retrieval of stored shape models are becoming increasingly more prominent in fields such as medical imaging, molecular biology and remote sensing. We present a novel framework that directly addresses the necessity for a rich and compressible shape representation, while simultaneously providing an accurate method to index stored shapes. The core idea is to represent point-set shapes as the square root of probability densities expanded in a wavelet basis. We then use this representation to develop a natural similarity metric that respects the geometry of these probability distributions, i.e. under the wavelet expansion, densities are points on a unit hypersphere and the distance between densities is given by the separating arc length. The process uses a linear assignment solver for non-rigid alignment between densities prior to matching; this has the connotation of “sliding” wavelet coefficients akin to the sliding block puzzle L'Âne Rouge. We illustrate the utility of this framework by matching shapes from the MPEG-7 data set and provide comparisons to other similarity measures, such as Euclidean distance shape distributions.

1. Introduction

Today’s scientific (and non-scientific) community generates information at a frantic pace; thus placing a paramount emphasis on developing flexible and robust systems for mining the data. Often, the desires are to classify test data, cluster similar groups or discover the closest match to an incoming query. The key enablers of these operations are the similarity metrics used for querying the data [1]. In this paper we focus on similarity metrics for shape models having applicability to a variety of disciplines, e.g. medical imaging, remote sensing and robotics. Our framework introduces a new shape representation and then uses the natural geometry arising from this representation to derive a geodesic-distance, similarity metric.

The present effort is motivated by a recent wavelet density estimation method [2] that estimates $\sqrt{p(x)}$ and then obtains a bona fide density as $(\sqrt{p(x)})^2$. This has several advantages over estimating $p(x)$ directly such as guaranteeing non-negativity and imposing a simple constraint on

the wavelet coefficients. This new density estimator uses a wavelet expansion of $\sqrt{p(x)}$, i.e.

$$\sqrt{p(x)} = \sum_{j_0,k} \alpha_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0,k} \beta_{j,k} \psi_{j,k}(x), \quad (1)$$

where $\alpha_{j_0,k}$ and $\beta_{j,k}$ are coefficients for the father $\phi(x)$ and mother $\psi(x)$ basis function; the j -index represents the current scale level and the k -index the integer translation value. (Note: $\phi(x)$ and $\psi(x)$ are also referred to as the scaling and wavelet functions respectively.) For numerical implementation, the infinite expansion in (1) is truncated to some n set of scale levels and we must also select a starting scale level j_0 . As discussed in [2], the coefficients in (1) are estimated with a maximum likelihood objective function which is minimized using a modified Newton’s method.

Expanding \sqrt{p} with a wavelet basis serves as the spring board to our development of an efficient similarity metric between shapes. We will show that given point-set shapes, we can use this density estimation method to represent shapes as probability densities—a natural by product is that the densities visually resemble the shapes. (For the purposes of this paper, we consider only two dimensional shapes but the theory and algorithmic procedures readily extend to higher dimensions.) All shapes in a given data set can be similarly represented. This representation has excellent properties like: (1) the multiscale wavelet coefficients of the densities can be thresholded [3] to compress the storage requirements (2) several different orthonormal bases can be used to estimate the densities, thus enhancing their descriptive capabilities and (3) the compact nature of wavelets provides both spatial and frequency localization enabling the densities to closely mimic shape features.

Based on this representation, the intuition for the similarity metric follows from considering the coefficients of the probability density $\{\alpha_{j_0,k}, \beta_{j,k}\}$ as the coordinates $c = [\alpha_{j_0,1}, \dots, \alpha_{j_0,m}, \beta_{j,1}, \dots, \beta_{j,m}]$ indexing the location of a density on a unit hypersphere; then the distance between two distributions p_1 and p_2 indexed by their coordinates c_1 and c_2 , respectively, is given by

$$d(p_1, p_2) = \cos^{-1}(c_1^T c_2). \quad (2)$$

(The unit hypersphere comes about from the constraints on the coefficients as discussed in §2.2.) We expand on these

intuitive ideas to develop a matching procedure that casts density matching in a linear assignment problem. The linear assignment is used to handle non-rigid differences between shapes. It “warps” the densities while preserving their defining properties, *e.g.* unit integrability and non-negativity. Since the densities closely resemble the shapes, we are in effect warping the shapes. It will be shown that this non-rigid alignment is necessary to obtain more accurate recognition. When one uses a Haar basis (box function) for the density expansion, the permutation of the wavelet coefficients due to the linear assignment visually looks like sliding blocks. Thus we have informally branded this process as *Shape L’Âne Rouge* after the French moniker for sliding block puzzles. Our method has several benefits, including:

- Shapes are not limited by topological constraints (such as the need to represent shapes as closed curves), eliminating extra effort often spent in developing parametrizations or other preprocessing.
- All of the intensive computations happen offline, *e.g.* the density estimation.
- For querying, the similarity metric computation between source and target shape is fast, satisfying the requirements for demanding indexing and retrieval applications.
- Use of wavelet representations enables flexibility in compression and storage.

1.1. Related Work

Existing work in shape modeling and matching span a broad spectrum of representations and their corresponding metrics. There are several recent surveys, *e.g.* [4], that succinctly describe shape representations such as unstructured point-sets or curves. They also detail the myriad of similarity measures that provide a means by which to compare shapes under a common representation. The advances made by all these methods have been instrumental in enabling robust indexing and retrieval mechanisms.

Because we incorporate a linear assignment solver to handle non-rigid deformations, our method is situated in close proximity to techniques that use transportation and assignment problem formulations [5] to obtain their distance measures. One such measure is the Earth Mover’s Distance (EMD) [6], a metric between general mass distributions of objects. Given two distributions \mathbf{x} and \mathbf{y} , the goal becomes to find a matrix $f_{i,j}$ that establishes a flow between all features x_i and y_j in \mathbf{x} and \mathbf{y} . Feasible flows must satisfy row sum, column sum and total sum constraints. Obtaining the flow and subsequently the EMD is generally based on the solution to the transportation problem [7]. Hence, one of

the main differences between our approach and EMD is that we solve a matching problem in contrast to the transportation problem. The EMD also requires one to decide on the features as well as the appropriate weighting of each feature per object. For some applications these choices may already be readily apparent, but for most this requires an added level of effort and investigation. Our method simply works on the point sets that naturally arise either from sampling or preprocessing.

Our present method also falls in the same paradigm as a recently introduced shape analysis framework [8] which uses geodesic distances on the manifold of Gaussian mixture models (GMMs) to establish a shape similarity metric. In this previous work, we represented shapes as mixture models and used the Fisher-Rao metric derived directly from the representation to obtain intrinsic distances on the manifold of parametric mixtures. Like this method, the present technique also leverages the geometry that results directly from the shape representation. However, when using GMMs it is not feasible to use the resulting metric for retrieval because the geodesics are not in closed-form. (GMMs present a large computational burden of solving for geodesic distances on arbitrary, high-dimensional manifolds.) With the present method, we have a well understood geometry with an easy to compute metric.

The remainder of this paper is organized as follows. In the next section we provide detailed discussions of our method—the representation of shapes as density functions expanded in a wavelet basis, the geometry that arises from this representation and the derivation of the similarity metric. We then follow with experimental verification of our method, Section 3. The indexing and retrieval accuracies are tested on a shape database consisting of 1400 shapes from the MPEG-7 Core Experiment CE-Shape-1 [9]. Our method is compared with another density-matching technique for retrieval: D2 shape distributions [10], for which we compute four different similarity measures. We also compare our results with published recognition rates of other algorithms on the MPEG-7 data. The last section concludes by summarizing our effort and proposing directions for future work.

2. Shape L’Âne Rouge

Our similarity metric, the geodesic distance on a unit hypersphere, is obtained directly from our representation of shapes as probability densities expanded in an orthonormal wavelet basis. This shape representation is detailed next, followed by a discussion on how this leads to the hypersphere geometry for the distributions. Afterwards, we illustrate the need for non-rigid alignment and how it can be accomplished on the space of distributions through a linear assignment formulation. It will turn out that the linear assignment process has to be regularized to improve matching

performance. To this end, we formulate a penalty term that restricts large movements of wavelet bases.

2.1. From Shapes to Wavelet Densities

The idea of representing shapes as densities is usually brought to fruition in two ways. Either the density is directly estimated from the shape’s discrete samples [11] or some other feature is first extracted from the shape and then the density is fit to these features [10, 6]; our method falls in line with the former. To our knowledge, this is the first time a *wavelet density estimator* has been used to directly represent shapes. Previous uses of wavelets in shape analysis [12] have been mainly restricted to extracting descriptors of contour shapes.

Many of the issues of estimating a bona fide density can be overcome by first estimating $\sqrt{p(x)}$ and then obtaining the desired density as $(\sqrt{p})^2$ [13, 14]. For two dimensional densities the wavelet expansion of the square root of the density is given by

$$\sqrt{p(\mathbf{x})} = \sum_{j_0, \mathbf{k}} \alpha_{j_0, \mathbf{k}} \phi_{j_0, \mathbf{k}}(\mathbf{x}) + \sum_{j \geq j_0, \mathbf{k}} \sum_{w=1}^3 \beta_{j, \mathbf{k}}^w \psi_{j, \mathbf{k}}^w(\mathbf{x}) \quad (3)$$

where $\mathbf{x} \in \mathbb{R}^2$, j_1 is some stopping scale level for the multi-scale decomposition and $(k_1, k_2) = \mathbf{k} \in \mathbb{Z}^2$ is a multi-index that represents the spatial location of the basis. (The translation range of \mathbf{k} can be computed from the span of the data and basis function support size.) The father and mother basis are tensor product combinations of their one dimensional counterparts, i.e.

$$\begin{aligned} \phi_{j_0, \mathbf{k}}(\mathbf{x}) &= 2^{j_0} \phi(2^{j_0} x_1 - k_1) \phi(2^{j_0} x_2 - k_2) \\ \psi_{j, \mathbf{k}}^1(\mathbf{x}) &= 2^j \phi(2^j x_1 - k_1) \psi(2^j x_2 - k_2) \\ \psi_{j, \mathbf{k}}^2(\mathbf{x}) &= 2^j \psi(2^j x_1 - k_1) \phi(2^j x_2 - k_2) \\ \psi_{j, \mathbf{k}}^3(\mathbf{x}) &= 2^j \psi(2^j x_1 - k_1) \psi(2^j x_2 - k_2). \end{aligned} \quad (4)$$

The goal is to estimate the set of coefficients $\{\alpha_{j_0, \mathbf{k}}, \beta_{j, \mathbf{k}}^w\}$ and reconstruct the density using (3). An efficient maximum likelihood method to estimate them, with fast convergence, is discussed in [2]. Due to the increased indexing notation for two dimensional wavelet expansion, we will typically resort to one dimensional arguments, as in Section 1, with it being understood that all results directly translate to two dimensions. Under a wavelet expansion of $\sqrt{p(x)}$, the unit integrability requirement of all probability densities translates to a constraint on the wavelet coefficients

$$\int \left(\sqrt{p(x)}\right)^2 dx = \sum_{j_0, \mathbf{k}} \alpha_{j_0, \mathbf{k}}^2 + \sum_{j \geq j_0, \mathbf{k}} \beta_{j, \mathbf{k}}^2 = 1. \quad (5)$$

Recall that we are using only orthonormal bases such as Haar, Coiflets or Symlets. Figure 1 illustrates estimated densities for four point set shapes, using a single level

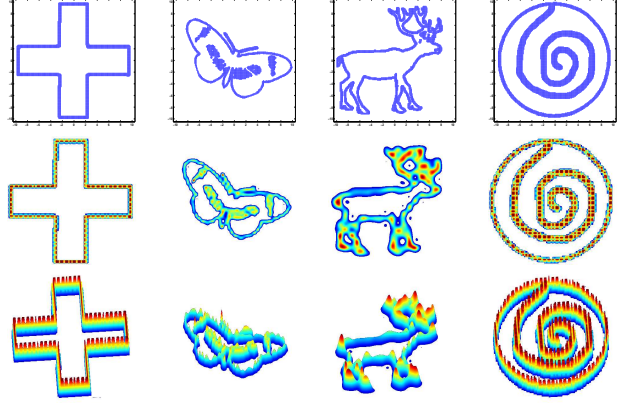


Figure 1. Example wavelet densities estimated from points-sets of MPEG-7 shapes. Top row are point sets, cardinality from left to right: 4,948;5,578;7,773;11,984. Second row is a nadir view of the estimated densities using the following wavelet families (from left to right): Haar ($j_0 = 2$), Coiflet-4 ($j_0 = 1$), Symlet-10 ($j_0 = 0$) and Haar ($j_0 = 2$). Third row is the perspective view. Notice how the wavelet densities accurately represent the shapes.

wavelet decomposition (with only scaling functions). The points were extracted from the MPEG-7 binary image data set. Notice how the compact nature of the bases does an excellent job in modeling the shape features. In the overhead views, it is readily apparent how closely the densities resemble the shapes. We feel this direct visual association of the density and the shape provides a nice advantage over trying to extract features and then fit the density to the features. Also, notice that shapes exhibit a variety of topological properties like interior structures and disconnected components.

2.2. The Geometry of Wavelet Densities: Geodesic Distances on the Hypersphere

Equation (5) showed that a natural by-product of working with the square root of the density and then expanding it with an orthonormal wavelet expansion was that it imposed a constraint on the basis coefficients; namely the sum of squared coefficient values must equal one. This immediately leads to the interpretation that the basis coefficients—which are unique to a particular density since wavelets serve as a true basis for the space of continuous distributions—give the coordinates for a position on the unit hypersphere. The ordering of the coefficients in the coordinate vector can be taken in any arrangement but it must be consistent across all densities. The dimensionality of the hypersphere is determined by the cardinality of the set containing all the coefficients. The hypersphere geometry of the densities can be more rigorously justified when we analyze the $\sqrt{p(x)}$ representation under the theoretical basis of information geometry [15, 16]. In this context, the Fisher information matrix (FIM) serves as the metric tensor on the manifold of a parametric family of distributions. One of the algebraic forms

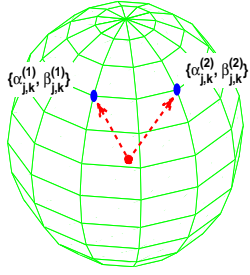


Figure 2. Hypersphere of densities. Unit integrability for densities requires $\sum_{j_0,k} \alpha_{j_0,k}^2 + \sum_{j \geq j_0,k} \beta_{j,k}^2 = 1$, also the FIM is reduced to the canonical metric of the unit hypersphere when \sqrt{p} is expanded in an orthonormal basis. This places the shapes represented by the densities on unit hypersphere with coordinates given by the wavelet coefficients. The above figure shows two densities, see coefficient superscript, on the hypersphere—their geodesic distance is the angle between the unit vectors.

of the FIM is given by

$$g_{u,v} = 4 \int \frac{\partial \sqrt{p(x|\Theta)}}{\partial \theta^u} \frac{\partial \sqrt{p(x|\Theta)}}{\partial \theta^v} dx \quad (6)$$

where $\Theta = \{\theta^1, \dots, \theta^m\}$ denotes the parameters of the distribution and u and v indicate the row and column index, i.e. for a family with m parameters the FIM is $m \times m$. Under an orthonormal expansion of $\sqrt{p(x|\Theta)}$, Eq. (6) reduces to the canonical metric tensor of a unit hypersphere embedded in an $m + 1$ Euclidean space. Rather than use the metric tensor to intrinsically compute geodesics on the hypersphere (an undertaking which would require us to parametrize the manifold), we can accomplish the same computation by realizing that the constraint $\sum_{i=1}^{m+1} (\theta^i)^2 = 1$ also implies the unit hypersphere geometry. Hence, closed-form geodesics distances can be simply computed using the usual angle measure between two unit vectors. Such is the case in our framework where $\sqrt{p(x|\Theta)}$ has been expanded in a orthonormal wavelet basis with the coefficients of the expansion serving as the parameters of the density, i.e. $\Theta = \{\alpha_{j_0,k}, \beta_{j,k}\}$. Two shapes represented as wavelet densities end up as two points on the hypersphere, see Figure 2. Since this is a unit hypersphere with the wavelet coefficients for each shape playing the role of two unit vectors, the angle between these unit vectors [Eq. (2)] immediately gives the geodesic distance between the shapes.

It is also interesting to note that we can obtain this *same* inner product interpretation required in (2) by taking the approach of working with a similarity measure directly between the densities, instead of analyzing the geometry implied by the coefficient constraints and the metric tensor. In particular, using the Hellinger divergence [17] to calculate the distance between two densities p_1 and p_2 gives

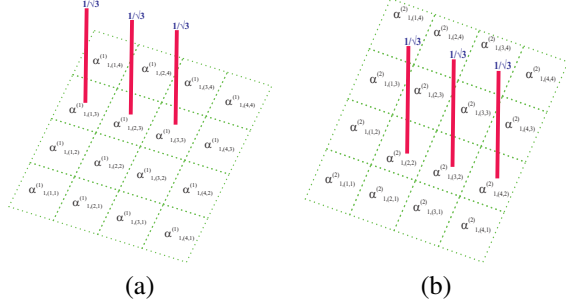
$$D_H(p_1, p_2) = \int_{\mathbb{R}^2} (\sqrt{p_1} - \sqrt{p_2})^2 dx \\ = 2 - 2 \left[\sum_{j_0,k} \alpha_{j_0,k}^{(1)} \alpha_{j_0,k}^{(2)} + \sum_{j \geq j_0,k} \beta_{j,k}^{(1)} \beta_{j,k}^{(2)} \right] \quad (7)$$

where $\{\alpha^{(1)}, \beta^{(1)}\}$ and $\{\alpha^{(2)}, \beta^{(2)}\}$ are the wavelet parameters of p_1 and p_2 respectively. Notice that we can factor out a -2 and drop the constant without effecting the qualities of the measure. This reduces (7) to an inner product between the coefficients of the densities, hence essentially giving the same measure as the one we derived above by analyzing the geometry of the space of distributions ($\cos^{-1}(\cdot)$ is not present). There are other notions of similarity measures between densities such as the Kullback-Leibler divergence and Euclidean distance but none of them operate on the square root of the density and they also do not provide a closed-form expression for the distance. We refer the reader to [6] for a summary of other distance measures between densities.

2.3. Sliding Wavelets

If our analysis ended with the previous section, we would be equipped with a very fast similarity metric. Given a pair of point-set shapes, we would merely estimate the wavelet coefficients of the square-root density of each shape and then take their inner product to get a measure of their closeness to each other. However, this approach is somewhat naïve in that it does not leverage the full mathematical formalisms that relate one shape to another. Following the Klein school of thought [18], similarity between shapes is often considered after quotienting out some transformation group, typically the group of similarity transformations [19]. Removing the transformations enables us to analyze effects that are intrinsic to the shapes. Non-rigid transformations are the most general, basically encompassing any continuous transformation. Practically it is expected that most shapes from the same category should differ by “smaller” non-rigid warps compared to shapes from other arbitrary categories; hence correcting for this prior to evaluating the similarity metric should enhance its discriminability. In our framework, we could incorporate non-rigid alignment in one of two ways: perform non-rigid alignment of the point sets prior to fitting the wavelet density or fit the density to the data and then adjust for non-rigid deformations by warping the densities. The former method usually involves adopting a spline based model to represent the non-rigid transformation [20] and can involve iterative optimization to solve for the spline parameters. Though these methods are able to model a large class of non-rigid deformations, they do not possess the computational efficiency needed for querying systems. Our method takes the second option of warping the densities which we accomplish by locally translating wavelet coefficients.

We now give a simple example to illustrate how warping the densities by local translations can increase recognition. Suppose two shapes have been affine aligned and there only remains a non-rigid warp between the two. We model the non-rigid deformation, in the infinitesimal, as local transla-



$$c_1 = \begin{bmatrix} 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$$

$$c_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & 0 \end{bmatrix}^T$$

Figure 3. Local non-rigid effects and the need for linear assignment. (a) is density p_1 of the first shape, with only scaling coefficients, $c_1 = [\alpha_{j,\mathbf{k}}^{(1)}]^T$, shown. (b) is the second shape with density p_2 with coefficients $c_2 = [\alpha_{j,\mathbf{k}}^{(2)}]^T$. Locally the point sets only differed by a translation which resulted in the densities differing by a translation. Without linear assignment the coefficient vectors of these would give an inner product of 0 and consequently large geodesic distance on the hypersphere. Linear assignment can correctly recover the local translation and then the geodesic distance will be small, reflecting the true similarity between the shapes.

tions. Figure 3 shows the estimated densities of two hypothetical shapes, see (a) and (b). The coefficients for the basis functions of each shape are indicated by a red bar. The density function shown in (a) only differs by a translation to density (b). Notice that if we were to stack the coefficients in a vector (from bottom left to top right) for each density and perform an inner product between them, the resulting value would be zero. This leads to high geodesic distance, $\cos^{-1}(0) = \frac{\pi}{2}$. However, if we simply slide the wavelet bases of one shape to align to locations on the other, our inner product would then yield a very high correlation indicating the true similarity between the shapes. Also we must be careful that whatever mechanism we use to translate the bases does not alter the values of their coefficients and compromise the properties of a bona fide density, i.e. (5) must hold to maintain unit integrability. The most straightforward way to accommodate these objectives is to reformulate our similarity metric under the action of a permutation group on the ordering of the coefficients. These specific requirements can be addressed within a linear assignment construct [5]; thus our deformation model can be interpreted as a “sliding grammar” wherein we only allow wavelets at each level j to independently slide to get a good match. The independent sliding assumption at each level implies that the “probability density mass” corresponding to each wavelet is independent of the rest. Consequently, this allows us to independently slide each wavelet to get a best match while maintaining the unit integrability constraint. While this justifies the independence assumption, “deformation grammars” more complex than sliding could

be considered, e.g. splitting coefficients. However, in this paper, we restricted ourselves to only sliding the wavelets leaving more exotic rules for future research. Even though each wavelet is allowed to slide, we cannot allow the sliding wavelets to collide and end up at the same spatial location. This imposes a permutation constraint on the sliding wavelets and the resulting deformation picture evokes the L’Âne Rouge puzzle, see Figure 4. Thus our new objective to minimize becomes

$$D(p_1, p_2; \pi) = -2 + 2 \left[\sum_{j_0, \mathbf{k}} \alpha_{j_0, \mathbf{k}}^{(1)} \alpha_{j_0, \pi(\mathbf{k})}^{(2)} + \sum_{j \geq j_0, \mathbf{k}} \beta_{j, \mathbf{k}}^{(1)} \beta_{j, \pi(\mathbf{k})}^{(2)} \right] \quad (8)$$

where $\pi(\mathbf{k})$ is a permutation operator that takes as input the wavelet spatial index \mathbf{k} and returns a new index \mathbf{k}' at the same level. (Since the wavelet coefficients can all be reversed to get the same density, there’s an overall sign symmetry which is accounted for in the linear assignment algorithm by running it twice—once with the set of coefficients $\{\alpha_{j_0, \mathbf{k}}, \beta_{j, \mathbf{k}}\}$ and a second time with $\{-\alpha_{j_0, \mathbf{k}}, -\beta_{j, \mathbf{k}}\}$.) The space of possible permutations is large and hence this objective needs to be regularized to yield useful results. Otherwise, every *source* shape’s coefficients could be re-ordered to be in the shape of the *target*; this is a detriment to recognition since any shape can essentially match another. To overcome this effect, we penalize large spatial movements by incorporating a cost based on the Euclidean distance between the centers of basis functions. This restricts large movements of the coefficients forcing them to be only locally translated. Incorporating this penalty gives our final objective function

$$E(\pi) = D(p_1, p_2; \pi) + \lambda \left[\sum_{j_0, \mathbf{k}} \|\mathbf{r}(j_0, \mathbf{k}) - \mathbf{r}(\pi(j_0, \mathbf{k}))\|^2 + \sum_{j, \mathbf{k}} \|\mathbf{r}(j, \mathbf{k}) - \mathbf{r}(\pi(j, \mathbf{k}))\|^2 \right] \quad (9)$$

where $\mathbf{r}(j, \mathbf{k})$ is a location operator—essentially giving us the center of the wavelet basis at (j, \mathbf{k}) —which has two inputs, the level j (and this includes j_0), the wavelet spatial index \mathbf{k} and returns a spatial location $\mathbf{r} \in \mathbb{R}^2$. The basic idea here is that as the regularization parameter λ is increased, the objective increasingly favors shorter wavelet sliding movements and hence smaller deformations. The optimal permutation π^* can be obtained by setting up the cost matrix

$$\mathcal{C} = c_1 c_2^T + \lambda d \quad (10)$$

where c_i is a vectorized representation of all the density wavelet coefficients for shape i and the matrix d contains pairwise distances between the wavelet basis locations. Figure 4 illustrates the effect of λ on the linear assignment and hence the similarity metric.

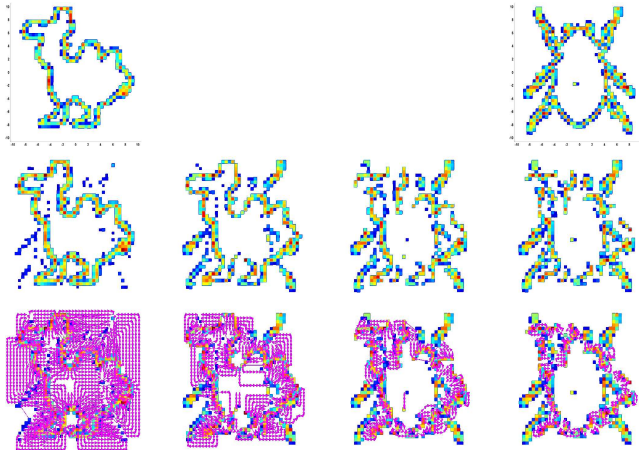


Figure 4. Effects of λ on linear assignment. Top row far left is target shape and far right is the source. Second row shows for small λ the source shape is almost perfectly transformed to the target while for large λ the source shape retains original shape; λ values from left to right: 10, 250, 500, and 1000. Third row illustrates the wavelet coefficients movement in row two (best viewed in color). The densities were estimated using the Haar family with $j_0 = 1$.

3. Experiments

The presented technique was evaluated on the MPEG-7 database [9]. The original data set consists of 70 different categories with 20 observations per category for a total of 1400 binary images. Each image consists of a single shape. One of the main strengths of our method is its accessibility and ease of use. The first part involves simply taking the data samples for each object and using them to estimate $\{\alpha_{j_0, \mathbf{k}}, \beta_{j, \mathbf{k}}^w\}$ for the wavelet expansion of \sqrt{p} . In the context of shape indexing this phase is completely offline, i.e. wavelet densities for the entire database can be estimated once and before the actual similarity computation takes place. Next, to compare two shapes, we first use the regularized linear assignment (9) to handle non-rigid effects and then use closed-form distance on unit hypersphere to obtain the similarity measure between them. We compare the performance of our method, Shape L'Âne Rouge, to D2 shape distributions [10].

For the MPEG-7 data, each shape was represented with a subset of points. There are *no topology or equal point-set cardinality requirements* amongst shapes, allowing shapes with richer features to be represented with a greater number of points, see Figure 1 for some examples. In this preliminary effort, we have focused on handling non-rigid effects. To this end, shapes within each category were affine aligned to a category reference shape. We used a recently introduced affine alignment algorithm that enables alignment of 2D point-set data without iterative optimization [21]. Once the shapes were aligned, all of them were brought into a common field of view by placing them in a $[-10, 10] \times [-10, 10]$ coordinate system. This was done

to control the translation range over which we estimate the densities. Next we estimated coefficients for the wavelet density of each shape using a Haar basis with $j_0 = 1$. Note it is possible to use several other families, but the Haar basis is available in closed form and reduces the time required to estimate the densities (on average about 2 to 3 minutes per shape). It is worth mentioning that regardless of the number of points used to represent each shape, once the densities are estimated all of them will have the same number of wavelet coefficients. (Recall that the densities are all estimated in the same square coordinate system.) Per these specifications, each wavelet density was represented with 1,764 coefficients.

Once the densities are estimated for all the shapes, pairwise matching between densities only involves working with the wavelet coefficients of the densities. When matching two shapes, the wavelet density coefficients of each are used to create the cost matrix in Eq. (10). With this cost matrix, we can then use the linear assignment solver presented in [22] to obtain the wavelet-coefficient rearrangements of the source shape with respect to the target. All of our experiments were conducted with multiple values of λ . For a shape pair, it typically takes less than 5 seconds to perform the linear assignment. Once the coefficients are re-ordered we can use Eq. (2) to obtain the geodesic distance between the shapes. In fact we experimented with three possible similarity measures that can be computed *after* the linear assignment: (1) the standard arc length geodesic distance after linear assignment, (2) geodesic distance plus the total distance penalty incurred for sliding and (3) just the total sliding penalty. (Note: The last two metrics are not be confused with the fact that the distance penalty is also used to regularize the sliding process which is different from treating the total amount of movement as a metric.)

We compared our method to D2 shape distributions as this is also a density-based shape retrieval metric. For each shape, a D2 shape distribution was created by taking 10,000 random pairwise distances between points on the shape. In [10], the authors then use these distances to construct a 1D histogram for each shape; this serves as a unique shape signature. Instead of using histograms, we estimate a 1D wavelet density for each shape. Distance metrics between shapes can be obtained by using a variety of 1D density dissimilarity measures. In addition to the Hellinger divergence, Eq. (7), we computed three other measures:

- Bhattacharyya: $D(p_1, p_2) = 1 - \int \sqrt{p_1 p_2} dx$
- χ^2 : $D(p_1, p_2) = \int \frac{(p_1 - p_2)^2}{p_1 + p_2}$
- L_2 : $D(p_1, p_2) = \left(\int (p_1 - p_2)^2 dx \right)^{\frac{1}{2}}$

Figure 5 shows some example D2 shape distributions using the 1D wavelet density estimator; these distributions correspond to shapes shown in Figure 1.

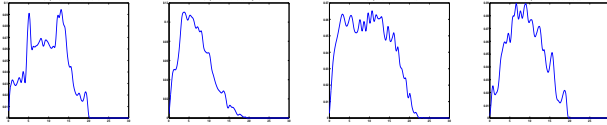


Figure 5. Example D2 shape distributions using wavelet densities estimators.. These distributions correspond to shapes in Figure 1 from left to right. All densities were estimated using a Symlet-7, $j_0 = 1$.

Shape L'Âne Rouge			D2 Shape Distributions	
Metrics	$\lambda = 500$	$\lambda = 2250$	Metrics	
Geodesic w/ LA	81.7%	85.25%	χ^2	59.3%
Geodesic + EDP	32.6%	12.1%	Hellinger	58.6%
EDP	32.5%	11.8%	Bhattacharyya	58.6%
			L_2	56.6%

Table 1. MPEG-7 recognition rate. Our method Shape L'Âne Rouge out performs D2 Shape Distributions [10]. In our method the choice of λ effects the recognition rate. See text for explanation of metrics. (LA \equiv linear assignment, EDP \equiv Euclidean distance penalty).

Performance on the MPEG-7 is most commonly evaluated using the bulls-eye criterion [9, 23]. Each shape is used as a query shape and the top 40 matches are retrieved from all 1400 shapes (the test shape is not removed). For a single query, maximum possible correct retrievals are 20 coinciding with the number of shapes in each category. Hence there are a total of 28,000 possible matches with the recognition rate reflecting the number of correct matches divided by this total. Table 1 lists the recognition rates using several density similarity measures for both Shape L'Âne Rouge and D2 shape distributions. Shape L'Âne Rouge significantly outperforms D2 shape distributions. This gives credence to the idea of working with feature representations that mimic the true visual properties of shapes, i.e. D2 shape distributions represent objects using a 1D signature derived from the 2D points whereas Shape L'Âne Rouge represents shapes using 2D densities which are visually similar to the shapes. The three different metrics computed for Shape L'Âne Rouge illustrate how λ impacts recognition performance. A judicious choice for λ can be made by optimizing over a training set. The different metrics also show that the wavelet density representation provides a rich set of features evident by that the fact the geodesic distance (with linear assignment) outperforms the metrics that include the Euclidean distance penalty. Hence the sliding alone is not sufficient to discriminate between shapes. (For high λ , the total sliding penalty dominates the second metric giving similar performance to the third.)

Recently, methods based on hierarchical representations [23, 24] have also reported recognition rates greater than 85% on the MPEG-7 data set. However, these methods work on a more simplified version of the problem than what we have addressed. They assume shapes are represented by their boundary outlines and typically use less than 200

points for the shapes. A hierarchical representation is used to capture both global and local properties. These methods have the drawback of extracting oriented, boundary curves which can be a troublesome preprocessing procedure. We also lose the descriptive power afforded by allowing arbitrary shape topologies and unconstrained point set cardinalities. The closest method, in terms of operating on unstructured point sets and not restricting shape topology, is [25] which has published recognition rate of 76.51% on the MPEG-7 data set. Our results clearly show reasonable gains over this method. We are still in the preliminary stages of exploiting the full capabilities of the Shape L'Âne Rouge framework, i.e. using multiscale representations to get more descriptive attributes, experimenting with different wavelet families, etc. Since we are already above 85%, we believe in the future these enhancements will improve our recognition rates significantly without sacrificing our ease-of-use and rich descriptive power.

4. Discussion

The development of robust and effective shape indexing and retrieval mechanisms largely depends on the representation model for the data and also the metrics used to distinguish one observation from another. In this paper, we have presented a novel shape representation scheme which gives rise to a natural metric that comes directly from the representation. Given an *unstructured point set model* of a shape, our Shape L'Âne Rouge framework estimates \sqrt{p} , under a wavelet expansion, directly from the point data and recovers the probability density as $(\sqrt{p})^2$. As we illustrated, these densities have a direct visual similarity to the original shape. The unit integrability property of all densities translates to a constraint on the wavelet coefficients, i.e. the sum squared coefficients equal one, see Eq. (5). Since the densities are uniquely identified by their wavelet coefficients, these are in effect the coordinates by which probability densities are indexed on a unit hypersphere. And since the densities represent the original shapes, intuitively the shapes are also on the unit hypersphere. As a result of this representation, we immediately gain a natural similarity measure between shapes by computing the arc length between probability densities on the unit hypersphere. Shape recognition can be improved if we adjust for non-rigid differences between a pair of shapes before computing a similarity measure. Rather than do this in the original shape space, we have introduced a novel way of deforming their wavelet density representation through the use of penalized linear assignment; allowing us to locally warp the density while maintaining its defining integrability and positivity properties.

Our framework has several advantages over other contemporary shape modeling and matching schemes:

- Each shape can have an arbitrary number of points

without topological restrictions. This is in sharp contrast to methods that work only on shape silhouettes or are limited to only a few sample points. Hence, the cardinality of a shape point set is dictated by the amount of points needed to accurately represent a shape's features and not by algorithmic limitations.

- Limited preprocessing is required since we directly take the shape points and estimate the density.
- The metric is in closed form and when incorporating linear assignment our method is still computationally efficient enough for querying applications.

We are still in the preliminary stages of fleshing out the capabilities of our Shape L'Âne Rouge technique. In the immediate future, we plan to incorporate the use of the multi-scale wavelet densities along with studying the effects of multiple wavelet families. We anticipate these will provide additional attributes for each shape which will further increase shape discriminability and subsequently improve recognition rates. We are also planning to investigate other penalty terms for the linear assignment objective function and better mechanisms for choosing λ .

Acknowledgments

This work is partially supported by NSF IIS-0307712 and NIH R01NS046812.

References

- [1] J. Tangelder and R. Velkamp, "A survey of content based 3D shape retrieval methods," in *Shape Modeling Applications*, 2004, pp. 145–156. [1](#)
- [2] A. Peter and A. Rangarajan, "Maximum likelihood wavelet density estimation with applications to image and shape matching," *IEEE Trans. Image Proc.*, vol. 17, no. 4, pp. 458–468, April 2008. [1](#), [3](#)
- [3] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, "Density estimation by wavelet thresholding," *Ann. Statist.*, vol. 24(2), pp. 508–539, 1996. [1](#)
- [4] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape benchmark," in *Proc. of the Shape Modeling Intl.* IEEE Computer Society, 2004, pp. 167–178. [2](#)
- [5] D. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1984. [2](#), [5](#)
- [6] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Intl. Journal of Comp. Vis.*, vol. 40, no. 2, pp. 99–121, 2000. [2](#), [3](#), [4](#)
- [7] F. Hitchcock, "The distribution of a product from several sources to numerous localities," *J. Math. Phys.*, vol. 20, pp. 224–230, 1941. [2](#)
- [8] A. Peter and A. Rangarajan, "Shape matching using the Fisher-Rao Riemannian metric: Unifying shape representation and deformation," *IEEE Intl. Symp. on Biomedical Imaging (ISBI)*, pp. 1164–1167, 2006. [2](#)
- [9] L. J. Latecki, R. Lakämper, and U. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *IEEE Conf. Comp. Vis. and Patt. Recog. (CVPR)*, 2000, pp. 424–429. [2](#), [3](#), [6](#), [7](#)
- [10] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Trans. on Graphics*, no. 4, pp. 807–832, 2004. [2](#), [3](#), [6](#), [7](#)
- [11] F. Wang, B. Vemuri, A. Rangarajan, I. Schmalfluss, and S. Eisenschenk, "Simultaneous nonrigid registration of multiple point sets and atlas construction," in *European Conference on Computer Vision (ECCV)*, 2006, pp. 551–563. [3](#)
- [12] G. C.-H. Chuang and C.-C. J. Kuo, "Wavelet descriptor of planar curves: theory and applications," *IEEE Trans. Image Proc.*, vol. 5, no. 1, pp. 56–70, January 1996. [3](#)
- [13] S. Penev and L. Dechevsky, "On non-negative wavelet-based density estimators," *Journal of Nonparametric Statistics*, vol. 7, pp. 365–394, 1997. [3](#)
- [14] A. Pinheiro and B. Vidakovic, "Estimating the square root of a density via compactly supported wavelets," *Comp. Stat. and Data Anal.*, vol. 25, no. 4, pp. 399–415, 1997. [3](#)
- [15] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*. American Mathematical Society, 2001. [3](#)
- [16] A. Srivastava, I. Jermyn, and S. Joshi, "Riemannian analysis of probability density functions with applications in vision," in *IEEE Conf. Comp. Vis. and Patt. Recog. (CVPR)*, 2007, pp. 1–8. [3](#)
- [17] R. Beran, "Minimum Hellinger distance estimates for parametric models," *Ann. Statist.*, vol. 5, no. 3, pp. 445–463, 1977. [4](#)
- [18] F. Klein, "A comparative review of recent researches in geometry," *Bull. New York Math Soc.*, vol. 2, pp. 215–249, 1892–1893, trans. M. W. Haskell. [4](#)
- [19] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. Wiley, 1998. [4](#)
- [20] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, June 1989. [4](#)
- [21] J. Ho, M. Yang, A. Rangarajan, and B. Vemuri, "A new affine registration algorithm for matching 2D point sets," in *Proceedings of the Eighth IEEE Workshop on Applications of Computer Vision*, 2007, p. 25. [6](#)
- [22] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, pp. 325–340, 1987. [6](#)
- [23] G. McNeill and S. Vijayakumar, "Hierarchical Procrustes matching for shape retrieval," in *IEEE Conf. Comp. Vis. and Patt. Recog. (CVPR)*, June 2006, pp. 885–894. [7](#)
- [24] P. Felzenszwalb and J. Schwartz, "Hierarchical matching of deformable shapes," in *IEEE Conf. Comp. Vis. and Patt. Recog. (CVPR)*, 2007, pp. 1–8. [7](#)
- [25] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002. [7](#)