# Robust Shape Retrieval Using Maximum Likelihood Theory

Naif Alajlan[1], Paul Fieguth[2], and Mohamed Kamel[1]

[1] PAMI Lab, E & CE Dept., UW, Waterloo, ON, N2L 3G1, Canada.
naif, mkamel@pami.uwaterloo.ca
[2] System Design Dept., UW, Waterloo, ON, N2L 3G1, Canada.
pfieguth@uwaterloo.ca

**Abstract.** The most commonly used shape similarity metrics are the sum of squared differences (*SSD*) and the sum of absolute differences (*SAD*). However, Maximum Likelihood (*ML*) theory allows us to relate the noise (differences between feature vectors) distribution more generally to a metric. In this paper, a shape is partitioned into tokens based on its concave regions, invariant moments are computed for each token, and token similarity is measured by a metric. Finally, a non-metric measure that employs heuristics is used to measure the shape similarity. The desirable property of this scheme is to mimic the human perception of shapes. We show that the *ML* metric outperforms the *SSD* and *SAD* metrics for token matching. Instead of the *ML* metric based on histograms for PDF approximation, which suffer from being sensitive to choices of bin width, we propose a Parzen windows method that is continuous and more robust.

## 1 Introduction

In recent years, content-based image retrieval (*CBIR*) has become a major research area due to the increasing number of generated images every day [1]. *CBIR* uses generic image features such as color, texture, and shape to interpret the content of images. In this work, we are interested in using shape descriptors in *CBIR*. Given a query image, we would try to obtain a list of images from a database of shape images, which are most similar to the query image. This problem can be solved in two stages. Firstly, a feature vector represents the shape information of the image. Then, a similarity measure computes the similarity between corresponding features of two images.

A desirable property of a similarity measure is that it should mimic the human perception of shapes. In fact, it has been verified that metric distances between feature points are not suited to model perceptual similarity between shapes [2]. This fact is illustrated in Fig. 1, where shapes $a$ and $b$ are similar, i.e., $d(a, b)$ is small. Similarly, $d(b, c)$ is small. Whereas shapes $a$ and $c$ are very different, i.e., $d(a, c)$ is large. So, $d(a, b) + d(b, c) < d(a, c)$, which violates the triangular inequality. Therefore, the perceptual distance measure is non-metric. On the other hand, a metric distance has the desirable properties, i.e., symmetry,

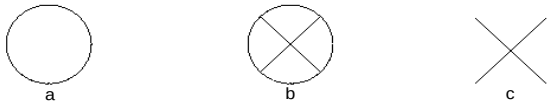



**Fig. 1.** Example of the triangular inequality failure. See text for explanation.

linearity, and triangularity, which make it efficient in determining the distance between two feature vectors.

In this paper we evaluate and compare shape retrieval efficiency using different metrics as the similarity measures for tokens. Each shape is partitioned into tokens in correspondence with its concave regions. Then, seven invariant moments are computed for each token, which are invariant to translation, scale, and rotation. A metric distance is used to measure the similarity between tokens. Three metric distances are considered, namely, *SSD*, *SAD*, and *ML*. A non-metric distance that employs heuristics is used to measure the similarity between shapes. It is chosen to be the majority vote, that is, a query shape is considered most similar to the shape in the database that shares the largest number of similar tokens with the query shape.

## 2 Shape Representation and Feature Extraction

Shape representation techniques can be categorized as structural versus global [3]. The main advantages of structural representations are the spatial localization of features and handling multi-object shapes. In the other hand, global representations are compact and, therefore, classical pattern recognition techniques can be applied. However, these global descriptors are imprecise to describe complex shapes. In order to take the advantages of both representations, a complex shape is decomposed into simpler shapes or tokens using a structural approach and global descriptors are obtained for the tokens.

### 2.1 Convex Hull

A non-convex shape can be analyzed by describing its concave regions. These can be identified by computing the difference between the convex hull of the shape and the shape itself. Borgefors and Baja proposed a technique to find the convex hull of a shape by repeatedly filling local concavities [4]. A good approximation of the convex hull can be achieved using $5 \times 5$ neighborhood of the shape's boundary elements. More precisely, the algorithm works as follows:

1. Each boundary pixel, i.e., a background pixel with at least one shape pixel neighbor, is labeled with the number of its shape pixel neighbors.
2. Boundary elements labeled more than 4, together with border elements labeled 4 and having at least one neighbor labeled more then 2 are changed to grey.

The above algorithm is repeated until all concavity regions are filled. The resulted grey envelope that includes the shape represents the convex hull as shown in the example of Fig. 2.

(a) (b)

**Fig. 2.** Result of the concavity filling algorithm: (a) original image, and, (b) the approximation of its convex hull.

### 2.2 Invariant Moments

The use of invariant moments to affine transformations (translation, scale, rotation, and skewness) is the most popular method for shape description. For a digital image, the moments are approximated by:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \tag{1}$$

Where the order of the moment is $(p+q)$, $x$ and $y$ are the pixel coordinates relative to some arbitrary standard origin, and $f(x,y)$ represents the pixel brightness. To make the moments invariant to translation, scale, and rotation, first the central moments are calculated:

$$\begin{aligned} \mu_{pq} &= \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \\ \bar{x} &= \tfrac{m_{10}}{m_{00}}, \bar{y} = \tfrac{m_{01}}{m_{00}} \end{aligned} \tag{2}$$

Then, the normalized central moments are computed:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\lambda}},\ s.t.\ \lambda = 1 + \frac{p+q}{2}\ \ and\ \ (p+q) \geq 2 \tag{3}$$

From these normalized parameters a set of invariant moments, found by Hu [5], can be calculated, which contain terms up to third order:

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \left( (\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right) + \\ &\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \left( 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right) \\ \phi_6 &= (\eta_{20} - \eta_{02}) \left( (\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \phi_7 &= (3\eta_{21} - \eta_{30})(\eta_{30} + \eta_{12}) \left( (\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right) - \\ &\quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \left( 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right) \end{aligned} \tag{4}$$

## 3 The Proposed Technique

An overview of the proposed system for shape retrieval is shown in Fig. 3. As can bee seen in the figure, two distinct measures of distance have been used: token (metric) and shape (non-metric) distances. The shape distance is obtained by combining token distances in order to derive a global measure of shape similarity. It is chosen to be the majority vote, that is, a query shape is considered most similar to the database shape that shares the largest number of similar tokens with the query shape. This measure is simple and, to some extent, mimics the human perception. Another desirable property of this scheme is that it provides means for partial matching.

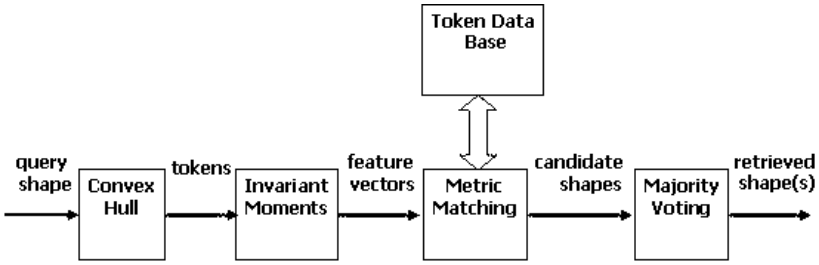



**Fig. 3.** The proposed system for shape retrieval.

All shapes in the database as well as a query shape are partitioned into tokens by subtracting each shape from its convex hull. To obtain a compact and discriminative description of these tokens, seven invariant moments are computed for each token. Then, for each query token, the database tokens are sorted using a metric. The metric matching in the tokens level results in a number of candidate shapes according to the user preference of the number of retrieved shapes. In the final stage, the candidate shape that shares the largest number of similar tokens with the query shape is considered the best match.

## 4 Maximum Likelihood Approach

In this section, three metrics that are used in the metric matching stage are viewed and explained. The difference vectors between each query token and all tokens in the database can be viewed as a noise with certain PDF. Sebe *et al.* showed how *ML* theory is used to relate the noise distribution to a metric [6]. Specifically, given the noise distribution, the metric that maximizes the similarity probability is:

$$\sum_{i=1}^{M} \rho\left(n_i\right) \tag{5}$$

Where $n_i$ represents the $i_{th}$ bin of the discretized noise distribution, $M$ is the number of bins, and $\rho$ is the maximum likelihood estimate of the negative logarithm of the probability density of the noise. In the case where the noise is Gaussian distributed, the PDF satisfies:

$$P\left(n_i\right) \propto \exp\left(-n_i^2\right) \tag{6}$$

Substituting (6) in (5) results in the so-called *SSD* or $L_2$ metric:

$$L_2(x,y) = -\sum_{i=1}^{M} \log\left(P\left(n_i\right)\right) = \sum_{i=1}^{M} \left(x_i - y_i\right)^2 \tag{7}$$

Similarly, for the two-sided exponential noise:

$$P\left(n_i\right) \propto \exp\left(-\left|n_i\right|\right) \tag{8}$$

Substituting (8) in (5) results in the so-called *SAD* or $L_1$ metric:

$$L_1(x,y) = \sum_{i=1}^{M} \left|x_i - y_i\right| \tag{9}$$

If the noise is Gaussian distributed, then (7) is equivalent to (5). Therefore, in this case the corresponding metric is *SSD*. In the same way, if the noise is exponential, then (9) is equivalent to (5) and the corresponding metric is *SAD*. However, if the noise distribution is neither Gaussian nor exponential, a metric can be extracted directly from the PDF of the noise, called the maximum likelihood metric, using (5):

$$L_{ML}(x,y) = -\sum_{i=1}^{M} \log\left(P\left(x_i - y_i\right)\right) \tag{10}$$

In practise, the probability density of the noise can be approximated as the normalized histogram of the differences between the corresponding feature vector elements [7]. For convenience, the histogram is made symmetric around zero by considering pairs of differences (e.g., $x-y$ and $y-x$). Nevertheless, the histogram approach for approximating the PDF of the noise suffers from being sensitive to the choice of the bin width (shift variant) and discontinuous. To overcome these drawbacks, Parzen windows method is employed where each noise point contributes linearly to the approximated PDF in the small proximity around that point using a given kernel function. Expressly, the approximated PDF is given by:

$$P(n) = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{h_n} \phi\left(\frac{n - n_i}{h_n}\right) \tag{11}$$

Where $M$ is the number of training points or kernels, $\phi(.)$ is the kernel function, and $h_n$ is the width of the kernel function. Too small width results in a noisy $P(n)$, where as too large width over-smoothes it.

## 5 Results and Discussions

In the following, the retrieval efficiency of the proposed system is evaluated with more emphasis on role of the metric matching of tokens on the overall performance of the system. Each training image in the database is partitioned into tokens based on its concavity regions and the seven invariant moments are computed for each token. The result is a token database of labeled feature vectors of fixed size. Then, the query image is partitioned in the same way as training images and each query token is matched to its closest tokens from the database using a metric. Three metrics are used for token matching, *SSD*, *SAD*, and *ML*. The similarity between two shapes is measured based on the largest number of the shared similar tokens.

A database of 216 images of 18 shapes (12 images per subject) is used to test our system as shown in Fig. 4. Two experiments are performed to evaluate the retrieval accuracy using different metrics. In the first, the aim is to test the system ability to retrieve the correct shape among a certain number of retrieved database shapes. In other words, the precision is plotted versus the number of retrieved shapes. The intention in the second experiment is to evaluate the system's ability to learn from few examples, i.e., the retrieval accuracy as the number of training images per subject changes.

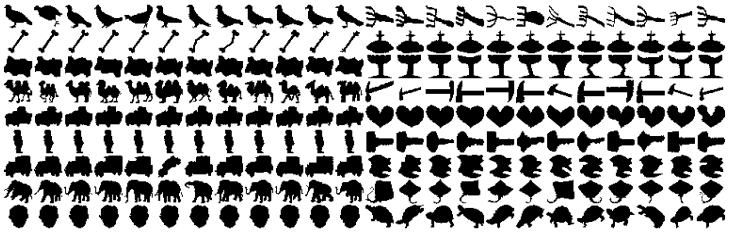


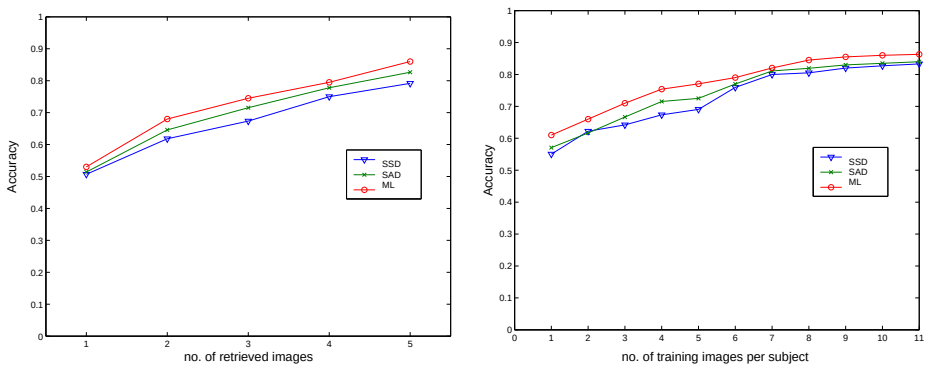
**Fig. 4.** The shape database used to test our system.

The results of the first experiment are shown in Fig. 5 (a). A retrieved shape is considered correct only if it belongs to the same subject of the query shape. So, similar shapes are not considered correct matches as long as they belong to different subjects. As can be seen in the figure, the *ML* metric outperforms both *SSD* and *SAD* metrics. Fig. 5 (b) shows the outcome of the second experiment. It can be noticed that the accuracy of the system does not improve significantly when more than five training images per subject are used, which means the system is able to learn from few examples. As in the first experiment, the *ML* metric does better than other metrics.

Finally, $\chi^2$ test is used to prove whether the noise distribution follows Gaussian or exponential distributions or not. For $\chi^2$ goodness-of-fit computation, the test statistic is defined as:

$$\chi^2 = \sum_{i=1}^{M} \frac{(O_i - E_i)^2}{E_i} \tag{12}$$

Where $O_i$ is the observed frequency and $E_i$ is the expected frequency for bin $i$. The hypothesis that the data are from a population with the specified distribution is rejected if $\chi^2 > \chi^2_{(\alpha,k)}$, where $\chi^2_{(\alpha,k)}$ is $\chi^2$ percentage point function with $k$ degrees of freedom and a significance level of $\alpha$. The results of $\chi^2$ test are shown in Fig. 6. It can be deduced that the noise distribution is not Gaussian nor exponential, although the exponential fit is better than the Gaussian. These findings justifies the outcomes of the previous experiments where the *SAD* metric performed better than *SSD* metric and the *ML* metric outperforms both *SSD* and *SAD* metrics.



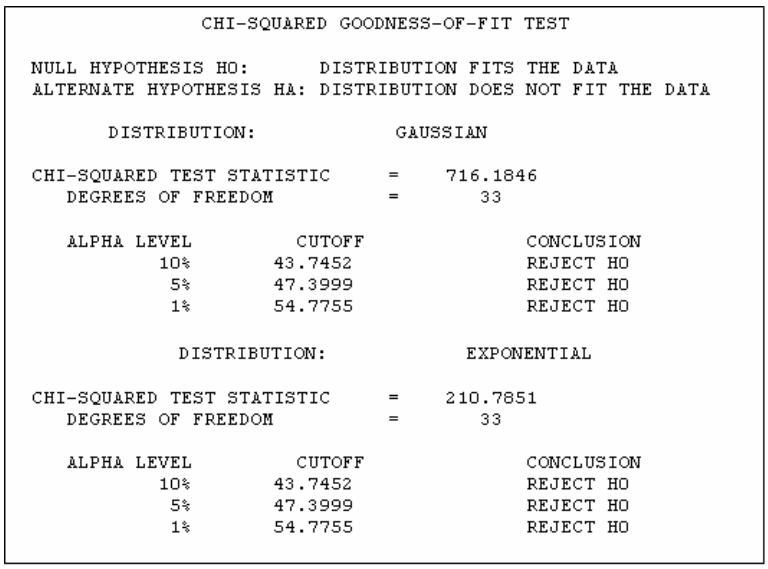

(a) (b)

**Fig. 5.** Accuracy of retrieval using different metrics versus (a) the number of retrieved images using one-third of the database for training and two-thirds for testing, and, (b) the number of training images per subject using three retrieved images.

## 6 Conclusions

In this paper, a scheme for shape retrieval is proposed. The used shape similarity measure is non-metric and, roughly, mimics the human perception of shapes. However, it makes use of the advantages of the metric similarity measures in the tokens matching. The problem of finding the appropriate metric to use for token matching is addressed. From the experiments, the *SSD* and *SAD* metrics are not justified because the similarity noise distribution is not Gaussian nor exponential, respectively. The *ML* metric, extracted directly from the noise PDF, outperformed both *SSD* and *SAD* metrics. Parzen windows method was used to approximate the noise PDF. It is more robust than the histogram method,

which is sensitive to the choice of the bin width. In the other hand, the main drawback of the *ML* metric is that, like most nonparametric approaches, it is computationally expensive. In applications where the speed is not a priority, the *ML* metric is a suitable choice.



```
                CHI-SQUARED GOODNESS-OF-FIT TEST

NULL HYPOTHESIS HO:      DISTRIBUTION FITS THE DATA
ALTERNATE HYPOTHESIS HA: DISTRIBUTION DOES NOT FIT THE DATA

       DISTRIBUTION:             GAUSSIAN

CHI-SQUARED TEST STATISTIC     =    716.1846
   DEGREES OF FREEDOM          =       33

   ALPHA LEVEL         CUTOFF              CONCLUSION
          10%        43.7452               REJECT HO
           5%        47.3999               REJECT HO
           1%        54.7755               REJECT HO

            DISTRIBUTION:            EXPONENTIAL

CHI-SQUARED TEST STATISTIC     =    210.7851
   DEGREES OF FREEDOM          =       33

   ALPHA LEVEL         CUTOFF              CONCLUSION
          10%        43.7452               REJECT HO
           5%        47.3999               REJECT HO
           1%        54.7755               REJECT HO
```

**Fig. 6.** Results of $\chi^2$ test of the goodness-of-fit of the noise to the Gaussian and exponential distributions.

## References


1. Berretti, S., Bimbo, A., Pala, P.: Retrieval by shape similarity with perceptual distance and effective indexing. IEEE Transactions on Multimedia. **2** (2000) 225–239
2. Berretti, S., Bimbo, A., Pala, P.: Retrieval by shape using multidimensional indexing structures. ICIAP. (1999)
3. Zhang, D.S., Lu, G.: Review of shape representation and description techniques. Pattern Recognition **37** (2004) 1–19
4. Borgefors, G., Sanniti di Baja, G.: Analyzing non-convex 2d and 3d patterns. CVIU. **63** (1996) 145–157
5. Hu, M.: Visual pattern recognition by moment invariants. IRE Trans. On Information Theory. **8** (1962) 179–187
6. Sebe, N., Lew, M.S., Huijsmans, D.P.: Toward improved ranking metrics. IEEE Trans. On PAMI. **22** (2000) 1132–1141
7. Sebe, N., Lew, M.S.: Maximum likelihood shape matching. ACCV. (2002) 713–718