

Shape Matching and Object Recognition Using Dissimilarity Measures with Hungarian Algorithm

¹D. Chitra, T. Manigandan, ³N. Devarajan

¹Government College of Technology Coimbatore, TamilNadu, India

²School of Electrical Sciences, Kongu Engineering College Erode, TamilNadu, India
manigandan_t@yahoo.com,

Abstract- The shape of an object is very important in object recognition. Shape matching is a challenging problem, especially when articulation and deformation of a part occur. These variations may be insignificant for human recognition but often cause a matching algorithm to give results that are inconsistent with our perception. In this paper, we propose an approach to measure similarity between shapes using dissimilarity measures with Hungarian algorithm. In our framework, the measurement of similarity is preceded by (1) forming the shapes from the images using canny edge detection (2) finding correspondence between shapes of the two images using Euclidean distance and cost matrix (3) reducing the cost by using bipartite graph matching with Hungarian algorithm. Corresponding points on two dissimilar shapes will have similar distance, enabling us to solve an optimal assignment problem using the correspondence points. Given the point correspondence, we estimate the transformation that best aligns the two shapes; regularized thin plate splines provide a flexible class of transformation maps for this purpose. The dissimilarity between the two shapes is computed as a sum of matching error between corresponding points, together with a term measuring the magnitude of the aligning transform. By using this matching error, we can classify different objects. Results are presented and compared with existing methods using MATLAB for MNIST hand written digits and MPEG7 images.

Keywords - Shape, Object Recognition, Euclidean, Hungarian Algorithm, and MPEG

I. INTRODUCTION

Object Recognition is a difficult task in the real world. Humans still out perform machines in most vision tasks, in both speed and quality. Our goal is to design machines that can recognize the objects at levels approaching or exceeding human performance. The shape of an object is very important in object recognition. Using the shape of an object for object recognition and image understanding is a growing topic in computer vision and multimedia processing [8], [11], [14].

Our objective in this paper is to operationalize the shape dissimilarity using Euclidean distance and hungarian for category-level recognition. There are three stages: 1) Forming the shapes from the images using canny edge detection. 2) Finding correspondence between shapes of the two images using Euclidean distance and cost matrix. 3) Reducing the cost by using bipartite graph matching with Hungarian algorithm.

On Growth and Form [17], Thompson observed that related, but not identical shapes can often be deformed into alignment using simple coordinate transformations. Serge Belongie et al [4] presented a novel approach to measure similarity between shapes and exploit it for object recognition. They introduced the shape context to solve the correspondence problem. The dissimilarity between the two shapes is computed as sum of matching errors between corresponding points, together with a term measuring the magnitude of the aligning transform. Nearest-neighbor classification framework is used to finding the stored prototype shape that is maximally similar to that in the image. Results were presented for silhouetted, trademarks, handwritten digits and the COIL data set.

The shape can be defined as the equivalence class under a group of transformations. Shape matching is identifying the shapes when two shapes are exactly the same. The statistician's definition of shapes addresses the problem of shape distance, assumes the correspondence that are known. Shape matching can identify without finding the correspondence by using the intensity-based technique. An extensive survey of shape matching [4] in a computer can broadly be classified into two approaches.

1. Brightness based method
2. Feature based method

1.1 Brightness Based Method

Brightness based or appearance based method is a complementary view to feature based methods. Instead of focusing on the shape this approach make direct use of gray values within the visible portion of the objects. The brightness information is used to find the correspondences and to align the gray scale values to compare brightness of two different shapes. The Fitting Hand Craft Model (FHC) presented by Yuille [23] suggests a flexible model to built invariance to certain kinds of transformations, but it suffers from the need of human designed templates and the sensitivity to initialization when searching via gradient descent.

Elastic graph matching developed by Lades et al [12] involves both the geometry and photometric features in the form of local descriptors based on the Dynamic Link Architecture, which is a Gaussian derivative. The performance is evaluated using statistical analysis. The alternative is the feature based method.

1.2 Feature Based Method

Feature based method is done by using the boundaries of silhouette images. It uses only the outer markings. Boundaries are conveniently represented by arc length. The dynamic programming approach is used, which is fast but the drawback is that it does not return the correspondences. The advantage of the algorithm is fast and invariant to several kinds of transformation including some articulation and occlusion. It uses edge detector for 2D image.

There are several approaches for shape recognition based spatial configurations of small number of key points. These models are used to vote for a model without solving the correspondence. The following techniques are based on the feature based method.

On growth and form model developed by Thomson [18] observed that related or similar but not identical shapes can often be formed into alignment using simple coordinate transformation. This work illustrates that the related shapes have the unique structure but it differs on the properties like scaling, rotation, shearing and projections. So it can't be used directly by applying the corresponding transformation.

Fischler and Elschlager [7] proposed mass spring model that minimizes the energy by using the dynamic programming technique. Thin Plate Spline (TPS) is an effective tool modeled by Bookstein [5] for modeling coordinate transformation in several computer vision applications. The technique from the machine learning for function approximation using Radial Basis Functions (RBF) is adapted to the task, which shows a significant improvement over the naive method. One drawback of this method is that it does not allow for principal warp analysis.

Cootes et al [6] describes a method for building models by learning patterns of variability from a training set of correctly annotated images, which can be used for image search in an iterative refinement algorithm that employed by Active Shape Models.

Shape Features and Tree Classifiers introduced by Amit et al [1] describe a very large family of binary features for two-dimensional shapes. Adopting the algorithm to a shape family is fully automatic, once the trained samples are provided. The best error rate for a single tree is 0.7 percent and the classification rate is 98 percent. The standard method for constructing tree is not practical because the feature set is virtually infinite. Patrice Y. Simard et al [13] proposed a memory based character recognition using a transformation invariant metric, (new distance measure). They tested the method on large handwritten character databases and MNIST. The error rate of this method is 2.5%.

Yann LeCun et al [21] proposed object recognition with gradient-based learning. This paper shows that it recognizes simple objects with handwritten characters using set of features and it gives that better performance by recognizing multiple objects without requiring explicit segmentation of the objects. Yoram Gdalyahu et al [22] proposed a flexible syntactic matching of curves to classify the silhouettes. The

proposed algorithm is applied to silhouettes and dissimilarity measured is computed. Based on the dissimilarity values, the image database are organized into shape categories.

Longin Jan Latecki et al [15] proposed a shape similarity measure based on correspondence of visual parts using silhouettes image database. They implemented the digital curve evolution to simplify the shapes, segmenting errors and digitizing noise. Shape matching procedure gives an intuitive shape correspondence and is stable with respect to noise distortions. Thomas B. Sebastian et al [17] presented a novel approach to finding a correspondence between two curves. The correspondence is based on a notion of an alignment curve. The optimal correspondence is found by an efficient dynamic programming method. The correspondence is shown for hand written character recognition, prototype formation and object recognition.

Evangelos Milios et al [3] proposed a shape-matching algorithm for deformed shapes based on dynamic programming. This algorithm is superior when compared to the traditional approaches of shape matching and retrieval such as Fourier descriptors and geometric and sequential moments. Robust shape similarity retrieval developed by Attalla and Siy [2] is a novel shape retrieval algorithm that can be used to match and recognize 2D objects.

Contour flexibility is a technique developed by Jianzhuang Liu [9] for shape matching in which the predominant problem is matching the shapes, when articulation and deformation occurs. Schwartz and Felzen [16] describe a new representation for 2D objects that captures shape information at multiple levels of resolution.

Vijayakumar and McNeill [20] introduce Hierarchical Procrustes Matching a segment-based shape matching algorithm [14] which avoids problems associated with global or local methods and performs well on shape retrieval test.

The drawbacks of the existing methods are that they do not return correspondences, suffer from the need for human designed templates and are applied to 2D images and limited 3D images. The proposed algorithm gives to improve the performance of existing methods.

II. SHAPE MATCHING WITH DISSIMILARITY MEASURE

The object can be treated as point set. The shape of an object is essentially captured by a finite subset of its points. A shape is represented by a discrete set of points sampled from the internal or external contours on the object. These can be obtained as location of pixels as found by an edge detection. These shapes should be matched with similar shapes from the reference shapes. Matching with shapes is used to find the best matching point on the test image from the reference image.

The proposed algorithm for matching with the shapes contains the following steps;

1. Preprocessing
2. Feature extraction
3. Finding correspondence
4. Applying transformation

5. Similarity measures and
6. Classifying images

2.1 Preprocessing

The preprocessing method is modifying the image for best matching to the reference image. Noise cancellation is the one of the preprocessing techniques. For noise cancellation a number of methods can be used like Median filter, Mean filter, Gaussian filter, etc. The Gaussian filter has been used for noise cancellation.

Gaussian filters [14] are a class of linear smoothing filters with the weight chosen according to the shape of the Gaussian function. The Gaussian smoothing filter is very good filtering for removing noise drawn from a normal distribution. The zero mean Gaussian function in one dimension is

$$g(x) = e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

When the Gaussian is spread parameter sigma determines the width of the Gaussian. For image processing, the zero mean two dimension discrete Gaussian function is

$$g(i, j) = e^{-\frac{(i^2 + j^2)}{2\sigma^2}} \quad (2)$$

2.2 Feature Extraction

The image should be compared for the unique features; the image has the features like number of pixel, width, length, edges and brightness of the image. The edges can be used as the unique features of the input and the reference image for its simplicity and effective of matching.

2.2.1 Edge Detection

The vision processing identifies the features in images that are relevant to estimate the structure and properties of objects in an image. Edges are one such feature. Edges typically occur on the boundary between the two different regions in an image. Edge detection is an algorithm that produces a set of edges from images.

2.2.1.1 Canny Edge Detection

The canny edge detector is used for the detection of the edges of the image [14]. The canny edge detector is the first derivative of a Gaussian and closely approximates the operator that optimizes the operator that optimizes the product of signal to noise ratio and localization. The canny edge detection algorithm is summarized by the following notation.

The image is denoted by $I[i, j]$. The result from convolving the image with the Gaussian smoothing filter using separable filtering is an array of smoothed data,

$$S[i, j] = G[i, j; \sigma] * I[i, j] \quad (3)$$

Where the σ is the spread of the Gaussian and controls the degree of smoothing. The gradient of smoothed array $S[i, j]$ can be computed using the 2*2 first-difference approximations

to produce two arrays. $P[i, j]$ and $Q[i, j]$ for the x and y partial derivatives.

$$P[i, j] \approx (S[i, j+1] - S[i, j] + S[i+1, j+1] - S[i+1, j]) / 2$$

$$Q[i, j] \approx (S[i, j] - S[i+1, j] + S[i, j+1] - S[i+1, j+1]) / 2$$

The edges are obtained by using the canny edge detection algorithm. The steps of the canny edge detection are:

1. Acquiring the image.
2. Smooth the image with a Gaussian filter.
3. Compute the gradient magnitude and the orientation by using finite difference approximations for the partial derivations.
4. Apply no maxima suppression to the gradient magnitude.
5. Use the double threshold algorithm to detect and link edges.

2.3 Finding Correspondence

Finding correspondence is measuring the dissimilarity between the reference image and the test image [4]. In finding correspondence the statistical method is applicable. Statistical method is used to find the distance between the reference image and the test image.

2.3.1 Distance Measure

The distance between the two pixels can defined in many different ways. To find the correspondence between the images, the Euclidean distance is

$$D_{\text{EUCLIDEAN}}([i_1, j_1], [i_2, j_2]) = \sqrt{(i_1 - i_2)^2 + (j_1 - j_2)^2} \quad (4)$$

The distance is calculated for the point p_i on the first image to the every point on the second image. The matching distance forms the cost matrix. To minimize the total cost the bipartite graph matching is used.

2.3.2 Bipartite Graph Matching

Given the set of costs C_{ij} between all pairs of points p_i on the first shape and q_j on the second shape, the cost of the matching should be minimized.

$$H(\pi) = \sum_i C(p_i, q_{\pi(i)}) \quad (5)$$

To minimize the cost matrix the Hungarian method is chosen. By this Hungarian method, the minimized cost can be found for the shapes which do not have the equal number of points on the both the shape.

Hungarian Method

The algorithm checks for the square matrix, (the no. of rows are not equal to the no. of columns), then a dummy row or a column is added to the cost elements. Then the smallest cost in each row of the cost matrix is found. The smallest cost element from each element in that row is subtracted and at least one zero in each row of this new matrix will be there which is called the first reduced cost matrix. In the reduced cost matrix, smallest element in each column is to be found and the smallest cost element from each element in that column is subtracted.

As a result at least one zero in each row and column of the second cost matrix will be there. And the optimal assignment is as follows: (i) The rows have to be examined successively until a row with exactly one zeros is found, (ii) The procedure is repeated for the columns of the reduced cost matrix. An optimal assignment is found. From this optimal solution, the cost is reduced. The matching error is calculated as the sum of the minimum cost in the each row and the image can be put in the different class using this matching error with suitable classifier design.

IV. TRANSFORMATIONS

To reduce the error, the transformations are applied to the image. The transformation can be defined as the changing the image from one form to another. An affine transformation [14] is a transformation that preserves co linearity (i.e., all points lying on a line initially still lie on a line after transformation) and ratios of distances (i.e., the midpoint of a line segment remains the midpoint after transformation). The transformation can perform shearing, scaling, rotation, translation and shifting. Hence the affine transformation can match the images which change at these parameters (rotation, scaling and shearing). The transformation transforms the image into different shifting, therefore this transformed image will reduce the error considerably.

IV. CLASSIFYING IMAGES

The images are classified into the different classes by using the matching error. The k-NN classifier is used to classify the images. In this classifier, the matching error is given with 10 centered classes and depending on the closeness of the matching error the image has classified.

V. RESULTS

5.1 Digit Recognition

The algorithm is tested with the MNIST database. The algorithm has been developed with the help of MATLAB and MNIST database servers for comparison of different methods of handwritten digit recognition. So the MNIST database is used for testing and training the algorithm. The MNIST database contains 60,000 handwritten digits in the training set and 10,000 handwritten digits in the test set. The time required is also a parameter to identify the algorithm efficiency. It can be measured by calculating the time required for identification of the single character. Time required is 2.1262sec. The memory requirements for the training images are 50 MB. The average error rate is 0.03516. The recognition rate of the hand written digit is 96.44%.

The following figures show that the input and processed output images.



Fig. 5.1. Input Images



Fig. 5.2. Noise Cancelled Images



Fig. 5.3. Edge Detected Images

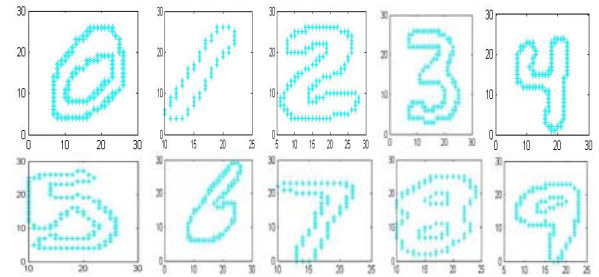


Fig. 5.4. Mapped Images

The input images, noise cancelled, edge detected and mapped images are shown in figures from 5.1 to 5.4. The sample images of the MNIST database are shown in the figure 5.5. The comparison of error rates for individual digits is shown in the figure 5.6. Table 1 shows the comparison of matching errors between shape context and Euclidean distance for sample MNIST handwritten digits.

TABLE I
 COMPARISON OF THE MATCHING ERROR WITH THE
 SHAPE CONTEXT AND EUCLIDEAN DISTANCE

Input Image	Reference Image	Shape context	Euclidean Distance
		2.85831	3.72801
		2.63122	3.21332
		2.93730	3.43440
		2.92876	3.32918
		2.75855	3.40228
		2.59482	3.41040
		2.32975	3.10104
		2.79757	3.26142

8	8	2.71720	3.59772
9	9	2.69774	3.23599



Fig. 5.5. MNIST database

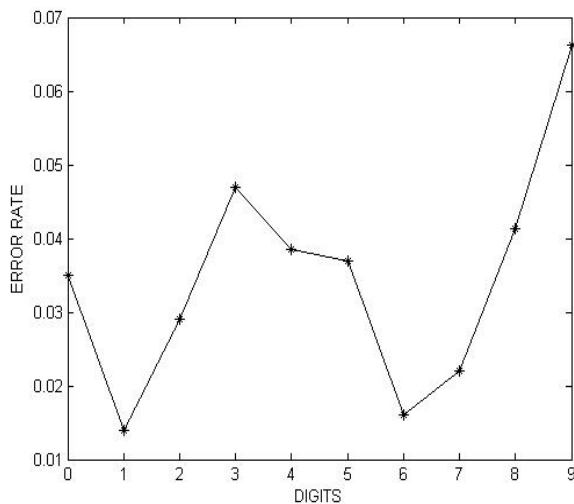


Fig. 5.6. Comparison of Error rates for MNIST database

5.2 MPEG-7 B Shape Database

The proposed technique is tested on the benchmark MPEG-7 B Shape database [9]. The database consists of 1,400 images, 70 shape categories, and 20 images per category. The sample images are shown in the figure 5.7. In a benchmarking test, each shape is taken as a query and 40 shapes with the highest scores are retrieved from the database. The task is repeated for each shape and the number of correct matches is noted. A perfect performance results in $1400 \times 20 = 28000$ matches. Results are given as a percentage of this perfect score. The table 2 shows the retrieval rate of the proposed technique using Euclidean distance with Hungarian algorithm. The retrieval rate for the MPEG-7 database is 89.55%.

TABLE 2
 RETRIEVAL RATES FOR MPEG-7 DATABASE

Algorithm	Score (%)
Shape context	76.51
Generative model	80.03
Curvature scale space	81.12
Chance probability function	82.69
Polygonal multiresolution	84.33
Multi scale representation	84.93
Inner distance	85.40
HPM-Fn	86.35
Shape Tree	87.70
Contour flexibility	89.31
Proposed method	89.55

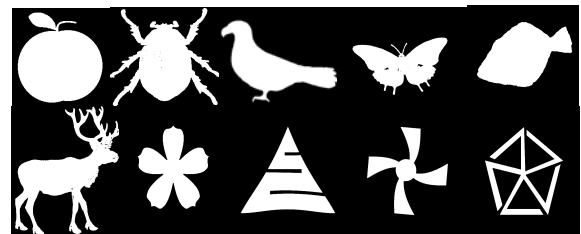


Fig. 5.7. MPEG-7 Shape database

6. CONCLUSION AND FUTURE SCOPE

This paper gives the efficient method for finding the class of the hand written images and MPEG-7 shape database. This method is simple, invariant to noise; CPU time is reduced, though the matching error is high. Retrieval rate of the databases is high compared to the previous methods. It can also be extended to the handwritten characters, industrial objects, face recognition and pedestrian identification.

Further, the algorithm will extend to recognize multiple objects from the images simultaneously. It can improve the reliability of the real time system. The algorithm can also be applied for the video image and aerial images to identify the object. The proposed method can be developed for different applications like image retrieval, military areas, investigation departments and industrial automation.

REFERENCES

- [1] Amit Y., Geman D. and Wilder K., "Joint Induction of Shape Features and Tree Classifiers", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 1300 – 1305, 1997.
- [2] Attalla E. and Siy P., "Robust Shape Similarity retrieval Based on Contour Segmentation Polygonal Multi resolution and Elastic Matching", *Pattern Recognition*, vol. 38 (12), pp. 2229 – 2241, 2005.
- [3] Evangelos Miliou and Euripides G.M. Petrakis, "Shape Retrieval Based on Dynamic Programming", *IEEE Transactions on Image Processing*, vol. 9, no.1, January 2000.
- [4] Belongie S., Malik J. and Puzicha J., "Shape Context: A new descriptor for shaping and object recognition", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 831 – 837, 2002.
- [5] Bookstein F.L., "Principal Warps: Thin-Plate Splines and Decomposition of Deformations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no 6, pp. 567 – 585, 1989.
- [6] Cootes T., Cooper D., Taylor C. and Graham J., "Active Shape Model - their training and Application", *Computer vision and image understanding*, vol. 61, pp. 38 – 59, 1995.
- [7] Fischler M. and Elschlager R., "The Representation and Matching of pictorial structures", *IEEE Transaction Computers*, vol. 22, no 1, pp. 67 – 92, 1973.
- [8] David Forsyth A., "Computer Vision" First Indian Edition, Pearson Education, 2003.
- [9] Jianzhuang Liu, "2D Shape Matching by Contour Flexibility", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 1 – 7, 2008.
- [10] Klassen E., Srivastava A., Mio W. and Joshi S.H., "Analysis of planar shapes using geodesic paths on shape Spaces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 26 (3), pp. 372 – 383, 2004.
- [11] Morton Nadler, 'Pattern Recognition Engineering', John Wiley & Sons, 2000.
- [12] Lades M, Vorbuggen C. and Lange J., "Distortion Invariant Object Recognition in the Dynamic Link Architecture", *IEEE Transactions on Computers*, vol. 42, no 3, pp. 300 – 311, 1993.
- [13] Patrice Y. Simard, Yann LeCun and John S. Denker, "Memory Based Character Recognition Using a Transformation Invariant Metric", *International Conference on Pattern Recognition*, vol. 2, pp. 262 – 267, IEEE Jerusalem, October 1994.
- [14] Rafael Gonzalez C., "Digital Image Processing", Second Edition, Prentice Hall Publication, 2006.
- [15] Longin Jan Latecki and Rolf Lakamper, "Shape Similarity Measure Based on Correspondence of Visual Parts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, no.10, October 2000.
- [16] Schwartz J., Felzen P.F., 'Hierarchical Matching of Deformable Shapes', *IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 1-8, 2007.
- [17] Thomas B. Sebastian, Philip N. Klein and Benjamin B. Kimia, "On Aligning Curves", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no.1, January 2003.
- [18] Thomson D.W., 'On Growth and Form', Cambridge Univ. press, 1917.
- [19] Tu Z. and Yuille A.L., 'Shape Matching and Recognition-Using Generative Models and Informative Features', *Proc. European Conf. on Computer Vision*, pp. 195 – 209, 2004.
- [20] Vijayakumar S. and McNeill G., "Hierarchical Procrustes Matching for Shape Retrieval", *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 885-894, 2006.
- [21] Yann LeCun, Patrick Haffner, Leon Bottou and Yoshua Bengio, "Object Recognition with Gradient-Based Learning", in Forsyth, D. (Eds), *Feature Grouping*, Springer, 1999.
- [22] Yoram Gdalyahu and Daphna Weinshall, "Flexible Syntactic Matching of Curves and Its Application to Automatic Hierarchical Classification of Silhouettes", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol.21, no.12, December 1999.
- [23] Yuille A., 'Deformable Templates for Face Recognition' *Journal of Cognitive Neuroscience*, vol. 3, no 1, pp. 59 – 71, 1991.