Multimedia Tools and Applications (2020) 79:4099–4114
https://doi.org/10.1007/s11042-019-07774-z

# Object recognition using Hausdorff distance for multimedia applications

K. Senthil Kumar [1] • T. Manigandan [2] • D. Chitra [2] • L. Murali [2]

Received: 1 November 2018 / Revised: 4 April 2019 / Accepted: 15 May 2019 /
Published online: 12 June 2019


## Abstract

The need for reliable and efficient systems for recognition of object from image is increasing day by day. A partial list of applications that may use such system includes searching and reading in hand written documents, recognizing digit on papers and others. In the existing work, Euclidean distance is used for recognizing object, but some of object it doesn't work well. The major aim of the work is to introduce new object recognition. So the proposed work recognizing object using a shape context and Hausdorff distance is introduced. The process analyses the layout of the image into digits. In the first step, the shape context is computed for two point set and Hungarian algorithm is used to find the correspondence between two point set. The process evaluates the similarity of the two point set using Hausdorff distance. Finally, the error rate is calculated by considering the affine cost and shape context cost. The algorithm tested using the MNIST, COIL data sets and a private collection of hand written digits and encouraging results were obtained. The error rate is reduced to 0.72%.

**Keywords** Object recognition · Shape matching · Hausdorff distance · Digit recognition · MNIST database

✉ K. Senthil Kumar
senthilasia@gmail.com

T. Manigandan
manigandan_t@yahoo.com

D. Chitra
chitrapacet@gmail.com

L. Murali
murlak37@gmail.com

[1] Tamilnadu College of Engineering, Coimbatore, India

[2] P.A. College of Engineering and Technology, Pollachi, India

## 1 Introduction

The field of object recognition has been researched for more than 40 years and impressive results were obtained. Unfortunately the hand written digit recognition still is the real challenges to researchers. The hand written digit is semi-cursive, where each digitcan be written in many form. Figure 1 shows the same digit 5 in different forms.

Shape analysis is a fundamental problem in computer vision and image processing, and affects a variety of application domains. It is important key aspect to understand the objects in the image. The geometry properties of the image provides robust signature for recognition under the noisy environment, and also the shape of the object has been used for matching the input image to the templates.

Shape can be represented by a set of points. It is an efficient method because of its low dimensionality [8]. Alternatively, the shape can also be represented as the regions enclosed by the planar curves [17]. It can be efficient under topological changes, and invariant to noise [7]. The efficient feature should be invariant under transformation. Since the shape is a property of geometric objects, it is invariant under transformation. The geometry properties of shape can be represented by a statistical function since its deals with the point set. It leads to the efficient analysis of geometric objects.

Li [14] discuss the detailed study of the invariants under various group of transformation in shape matching. Forsyth et al. [9] uses algebraic entities such as lines or polynomial curves for constructing primitive invariants based on a global shape descriptor. These descriptors takes longer time to complete the process.

Belongieet al [5] proposed a novel shape descriptor, called shape context which represents the shape explicitly with a radial histogram. The algorithm checked with MNIST, coil database and MPEG data bases and shows the better recognition rate. Attneave [3] represents the shape by curvatures as a shape descriptor due to its invariant properties and computational convenience. The undesirable parameterization of shape matching is avoided by using signature that consists of curvature and its first derivative [6]. Pedro F. Felzenszwalb [11] et al. described an object detection system based on mixtures of multiscale deformable part models. The system depends on discriminative training of classifiers. It is one of the efficient methods for matching deformable models to image.

The signature based differential invariants [10] provide the locality which is important character for identified occlusions. However, the main problem of differential invariant is the noise effect is also amplified. This problem may overcome by the scale space approaches via

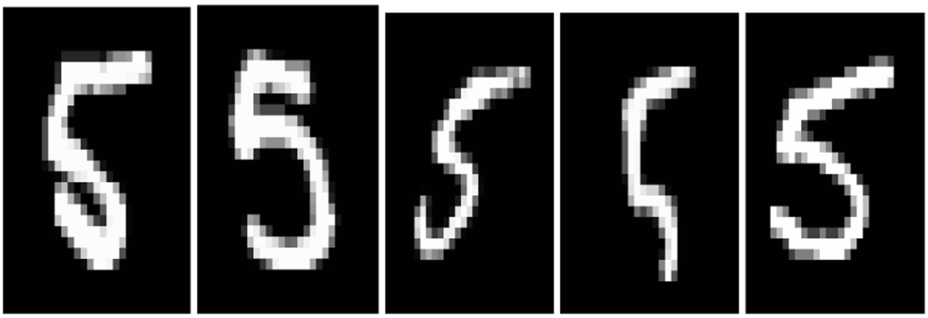


**Fig. 1** Example of hand written digits

diffusion process [19]. The existing algorithms are struggled with common digits like six and nine and also had the time consuming process. In the existing work, Euclidean distance is used for recognizing object, but some of object it doesn't work well.

Proposed work recognizing object using a shape context and Hausdorff distance is introduced. In the first step, the shape context is computed for two point set and Hungarian algorithm is used to find the correspondence between two point set. The process evaluates the similarity of the two point set using Hausdorff distance. The algorithm tested using the MNIST, COIL data sets and a private collection of hand written digits and encouraging results were obtained.

## 2 Proposed methodology

The process analyses the layout of the image into digits. In the first step, the shape context is computed for two point set and Hungarian algorithm is used to find the correspondence between two point set. The process evaluates the similarity of the two point set using Hausdorff distance. Finally, the error rate is calculated by considering the affine cost and shape context cost. The proposed algorithm for matching with the shapes contains Preprocessing, Feature extraction, Applying transformation, Finding correspondence, Similarity measures and 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. The Gaussian filter has been used for noise cancellation. Gaussian filters are a class of linear smoothing filters with the weight chosen according to the shape of the Gaussian function. The Fig. 2 shows the edge detected image of digit 2. 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

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

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

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

The object can be treated as point set. The shape of an object is essentially captured by a finite subset of points. A shape is represented by a discrete set of points sampled from the internal or external contours of the object. Contours can be obtained as location of pixels as found by an edge detection. The shapes should be matched with similar shapes from the reference shapes [2]. Matching [1, 18] with shapes is used to find the best matching point on the test image from the reference image.

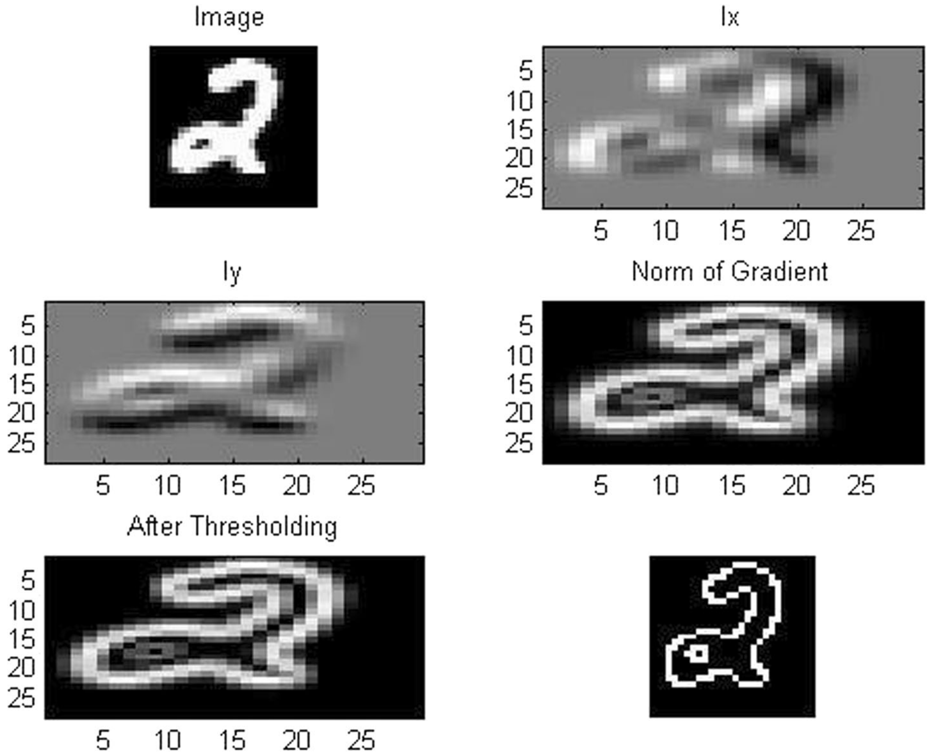



**Fig. 2** Input and edge detected images

### 2.2 Feature extraction

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

The canny edge detector [4, 15, 16] is used for the detection of the edges of the image. 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.

Detection of edge by means of lesser error rate, which means with the purpose of the detection must correctly catch as several edges shown in the image as likely.

The edge point detected from the operator must correctly localize on the center of the edge.

A given edge in the image must only be marked once, and where potential, image noise must not generate false edges.

### 2.3 Modeling transformation

The transformation is applied on shape to reduce the dissimilarity in matching. The transformation is in different forms for a given finite set of correspondences between points on two shapes; one can proceed to estimate a plane transformation Y: $R^2 \rightarrow R^2$ that may be used to map arbitrary points from one shape to the other. The affine transformation from a point x Є $R^2$ to the point y Є $R^2$as in Eq. (3)

$$Y = Ax + c \tag{3}$$

where A is m × m, c is m × 1 and A is non singular. A linear transformation is affine transformation of equation but with c = 0. For some matrix A (m × m) and a translational offset vector c (m × 1) parameterizing the set of all allowed transformations. The least squares solution Y = (A, c) is obtained by Eqs. (4) and (5)

$$c = \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{p_i}-\mathrm{q}_{\pi(\mathrm{i})}\right) \tag{4}$$

$$A = Q + P^t \tag{5}$$

where P and Q contain the homogeneous coordinates of p and q, respectively, i.e.,

$$P = \begin{pmatrix} 1 & p_{11} & p_{12} \\ \vdots & \ddots & \vdots \\ 1 & p_{n1} & p_{n2} \end{pmatrix} \tag{6}$$

Here $Q^+$ denotes the pseudo inverse of Q.

Thin plate spline [14] is a natural interpolating function for two dimensions and plays a similar role in the m = 2 dimensions to the natural cubic spline for interpolation in one dimension case. The natural cubic spline in one dimension is unique interpolate g(x) which minimizes the roughness penalty as in Eq. (7)

$$\int\left(\frac{\partial^2 g}{\partial x^2}\right)dx \tag{7}$$

subject to interpolation at the knots. For shape analysis, consider the (2 × 1) landmarks $t_j$, j = (1,…..,k) on the first shape mapped exactly into the $y_i$, i = (1,……,k) on the second shape, i.e. there are 2 k interpolate constraints, the Eq. (10) shows the interpolation between the landmarks of first shape $t_j$ and the second shape $y_j$

$$\left(y_j\right)_r = \varnothing_r\left(t_j\right) \tag{8}$$

Where r = 1,2; j = 1,…,k, $\varnothing_r(t_j) = (\varnothing_1\ (t_j\ )\ \varnothing_2\ (t_j\ ))^{T.}$

also written as,

$$\varnothing_r\left(t_j\right) = c + At + w^t s(t) \tag{9}$$

It can be proved that the transformation Eq. (12) minimizes the total bending energy (10) and (11) of all possible interpolating functions mapping from T to Y, where the total bending energy is given by,

$$\mathrm{I}\left(\varnothing_\mathrm{j}\right) = \iint'_{\mathrm{R}^2}\frac{\partial^2\varnothing_\mathrm{j}}{\partial^2\mathrm{x}} + \frac{\partial^2\varnothing_\mathrm{j}}{\partial\mathrm{x}\partial\mathrm{y}} + \frac{\partial^2\varnothing_\mathrm{j}}{\partial^2\mathrm{y}}\mathrm{dx\ dy} \tag{10}$$

$$I(\varnothing_j) = W^T sW \tag{11}$$

and has the form

$$\varnothing(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} w_i U\left(\|(x_i, y_i) - (x, y)\|\right) \tag{12}$$

Where the kernel function U(r) is defined by $U(r) = r^2 \log r^2$ and $U(0) = 0$ as usual. Here the r is Euclidian distance between the point set, In order for $\phi(x, y)$ to have square integral second derivatives, it is required that

$$\sum_{i=0}^{n} w_i = 0 \text{ and } \sum_{i=0}^{n} w_i x_i = \sum_{i=0}^{n} w_i y_i = 0 \tag{13}$$

Together with the interpolation conditions, $(x_i, y_i) = v_i$, this yields a linear system for the TPS coefficients as shown in Eq. (14).

$$\begin{pmatrix} K & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} w \\ a \end{pmatrix} = \begin{pmatrix} v \\ 0 \end{pmatrix} \tag{14}$$

where $K_{ij} = U(\|(x_i,y_i)-(x_j,y_j)\|)$, the ith row of P is $(1,x_i,y_i)$, w and v are column vectors formed from $w_i$ and $v_i$, respectively, and a is the column vector with elements $a_1,a_x,a_y$.

The proposed algorithm is an iterative based approach, the image is treated as point set, and this point set is matched using synthetic matching. The two sample images and the corresponding point sets is shown in Fig. 3.

## 2.4 Correspondence

Correspondence is found for the two point set by using the shape descriptor, shape context and shortest path augmenting algorithm. Figure 4 shows the corresponding points for the first shape to the second shape at first and fifth iteration. Further, with this correspondence, the aligning transform is applied on the second shape with reference to the first shape. The iteration continues until the best matching occurs or maximum iteration reaches. In this algorithm, the maximum five iterations processed for the best matching.

## 2.5 Similarity measures

The edges are taken as the feature of the object, the point set over the edges is considered as a shape. Shape of an object contains the group of points and the point may have two data set for the 2-d coordinates. This point set is considered for shape matching. The shape matching has been done using the hausdorff distance [12, 20].

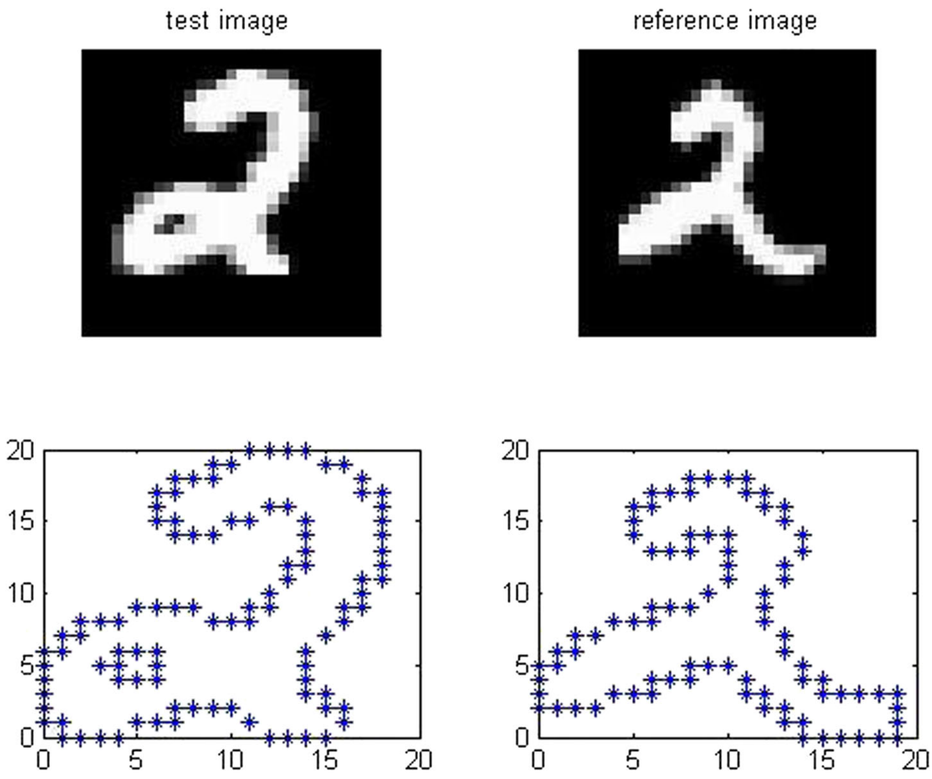



**Fig. 3** Test image and reference image with its point set

### 2.5.1 Hausdorff distance

Hausdorff distance from set A to set B is a maximin function, defined as The directed Hausdorff distance function

$$h(A,B) = \max_{a\in A}\min_{b\in B}(d(a,b)) \tag{15}$$

Where a and b are points of sets A and B respectively, and d(a, b) is any metric between these points. A more general definition of Hausdorff distance would be

$$\mathrm{H}\ (\mathrm{A},\mathrm{B}) = \max\ \{\mathrm{h}\ (\mathrm{A},\mathrm{B}),\mathrm{h}\ (\mathrm{B},\mathrm{A})\} \tag{16}$$

The two distances h(A, B) and h(B, A) are sometimes termed as forward and backward Hausdorff distances of point set A to point set B. The hausdorff distance used to calculate the distance from the test image and the reference images in the different class. The distance gives the similarity information of test shape to the shape in the different classes. The Fig. 4 shows the one to one matching between the test image and reference image. It shows the hausdorff distance is 1.4371.

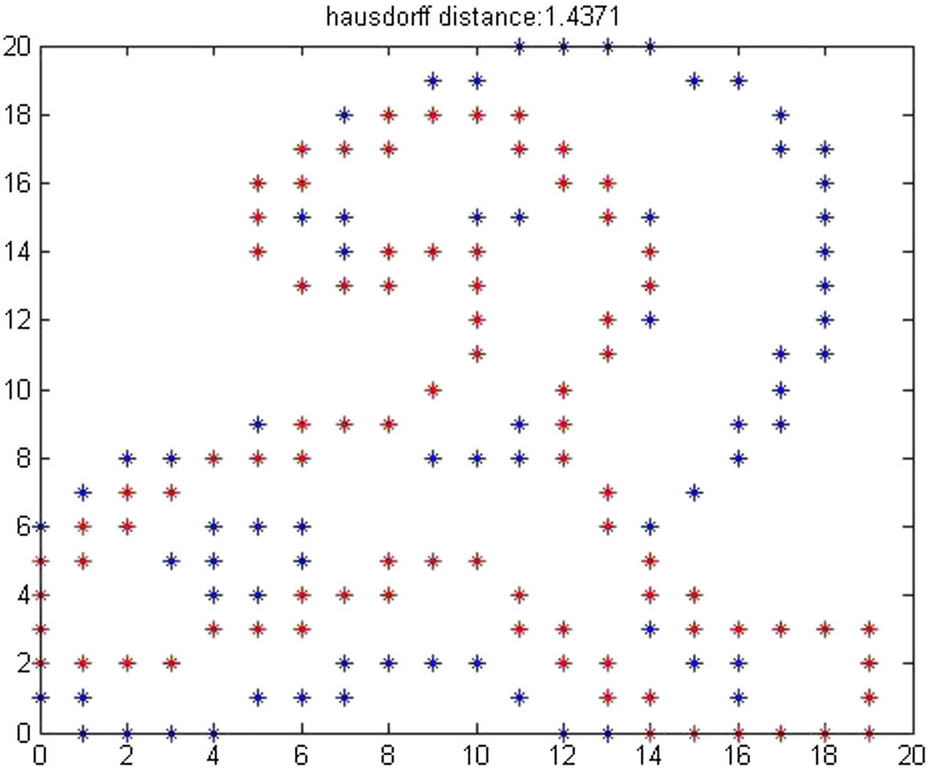



**Fig. 4** One to one matching between test image and reference image

## 2.6 Nearest neighboring classification

The NN classification is used to classify the test digits into the different classes with ten number of input layer, 15 number of hidden layer and 10 number of output layer. The principle behind nearest neighbor methods is to find a predefined number of training samples closest in distance to the new point, and predict the label from these. The number of samples in proposed method is two for each digit. The hausdorff distance is used as metric measures.

## 2.7 Proposed algorithm for object recognition

The following algorithm shows the step by step procedure for the shape matching using hausdorff distance. Figure 5 shows the flow chart for shape matching.

STEP 1 : Acquiring the images.
STEP 2 : Removing the noises from image.
STEP 3 : Finding the edges on the image and it can be treated as shape of the object.
STEP 4 : Matching the test shape and reference shapes from the data base using the hausdorff distance.
STEP 5 : Based on the distance the shapes are classified into the object.

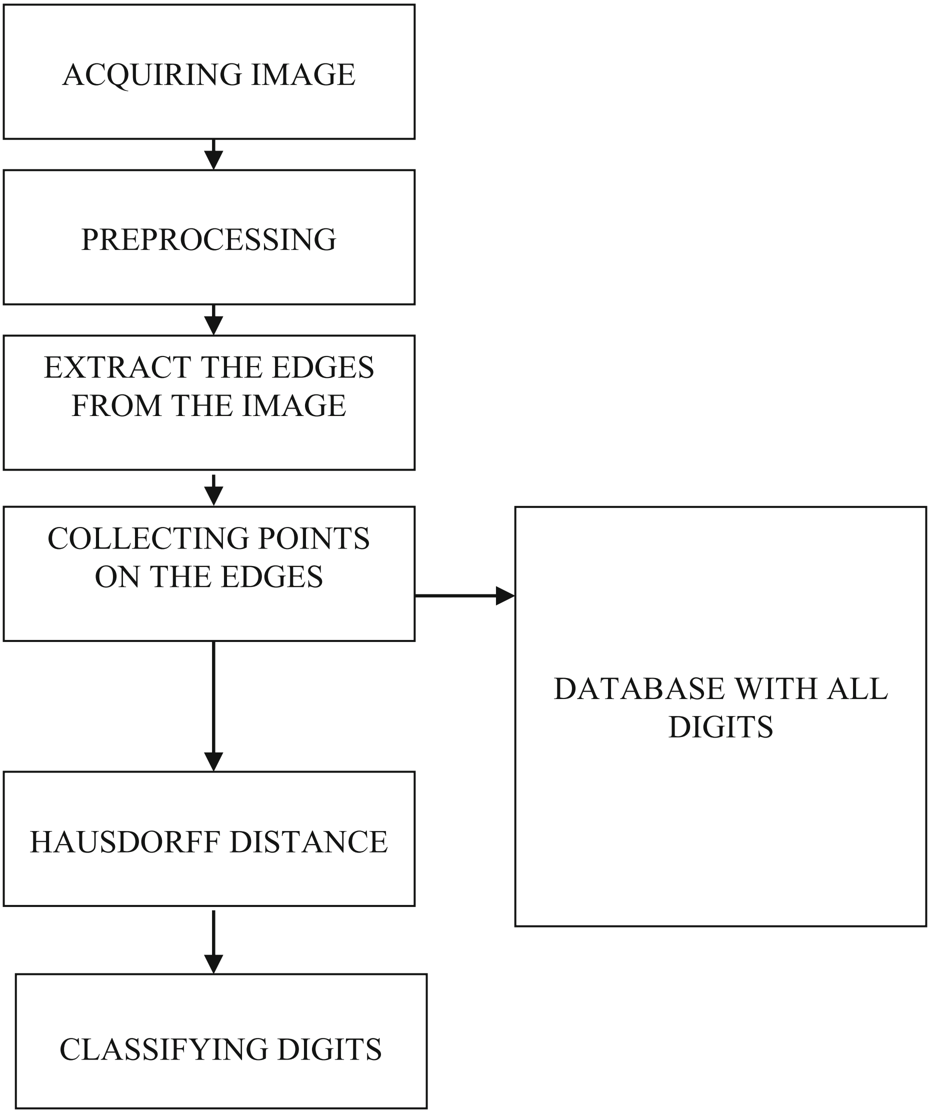



**Fig. 5** Flow chart of shape matching with hausdorff distance

## 3 Results

The algorithm is checked with the MNIST and COIL database. The MNIST database consists of the 2-D handwritten image. The edges and boundaries are easily identified, so 100 points from edges are taken for shape matching. While in the COIL database the 3-D images were used so, the internal and external edges were taken for shape matching. Three hundred edge points were considered for the COIL database shape matching.

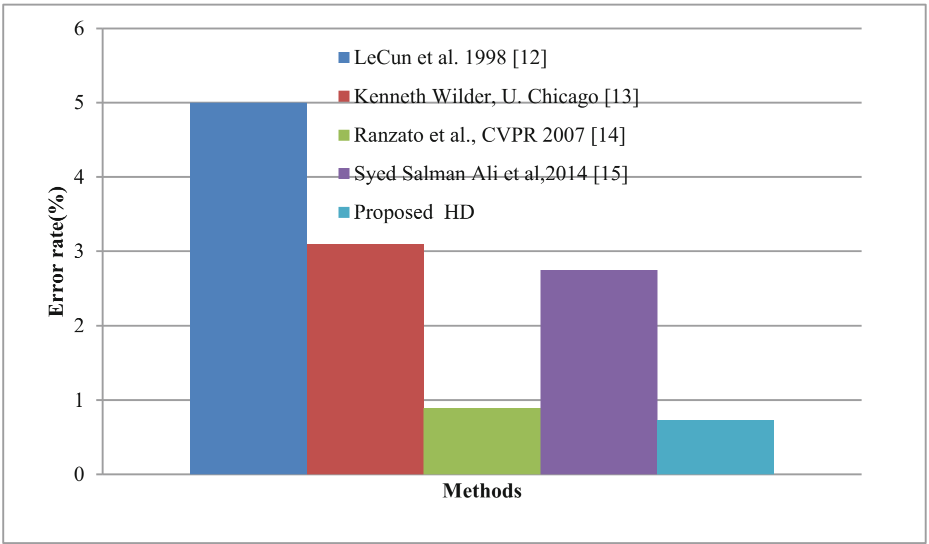



**Fig. 6** Error rate comparisons vs. recognition methods

### 3.1 MNIST database

The algorithm is checked with the MNIST database. The MNIST database consists of 60,000 training image and 10,000 test images. The database in the single data file, the hand written images are stored with the size of 784 in single row which has the image size of 28 × 28.The Fig. 6 shows the misclassified images of MNIST data set. The algorithm gives the reasonable results compare to the existing methods. The Algorithm recognizes 9968 images correctly out of 10,000 images from MNIST dataset. The error rate is reduced to 0.72%. The program is run with MATLAB 14b on 2GB RAM computer. The time required for single match is1.47 s. The Table 1 shows the comparison of error rate with the existing and proposed method shown in Fig. 6.

Syed Salman Ali et al. developed the system with DCT and HMM though the speed of the system increases the error rate reduced [1]. Since the proposed method yields better result compared to other methods, it may be extended for other dissimilarity measures. The Fig. 7 shows the no of error images with each digit and it shows the digit 4 has the more errors and digit 3 and 0 has lowest error graphically shown in Fig. 8.

**Table 1** Error rate comparisons

| Method | Reference | Error rate (%) |
|---|---|---|
| K-nearest-neighbors, Euclidean (L2) | LeCun et al. 1998 [13] | 5.0 |
| K-nearest-neighbors, Euclidean (L2) | Kenneth Wilder, U. Chicago [2] | 3.09 |
| large conv. net, random features [no distortions] | Ranzato et al., CVPR 2007 [18] | 0.89 |
| Recognition using DCT and HMMs | Syed Salman Ali et al. 2014 [1] | 2.74 |
| Proposed method | – | 0.73 |


**Fig. 7** Error digit in MNIST database

### 3.2 COIL database

The algorithm is applied for the COIL database. In COIL-20 database contains the images of real world objects in the different angle. The sampling point on the image is increased from 100 to 300 points, both the internal and external contour are taken into the consideration. The prototypes are included in the algorithms. The prototype selection is an important factor that

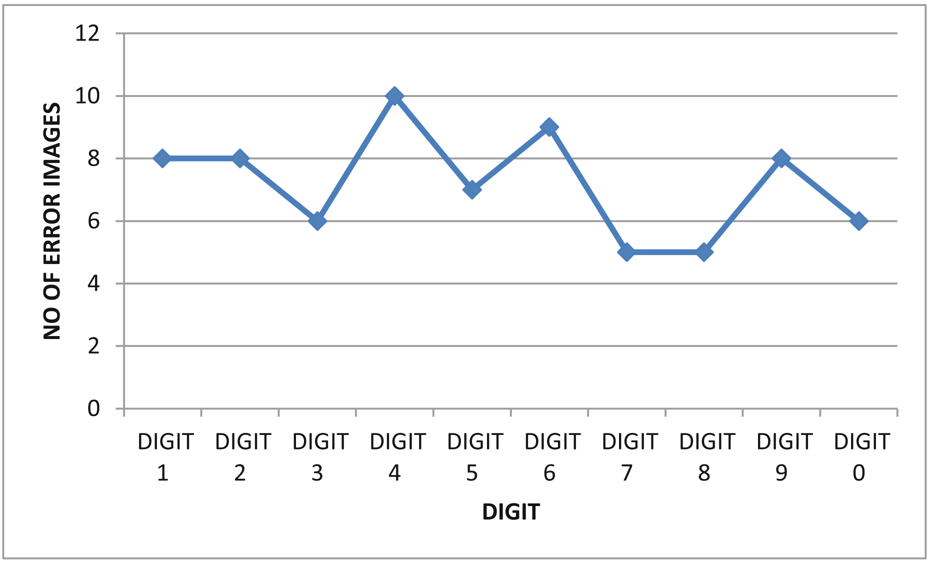



**Fig. 8** Number of error images for each digit

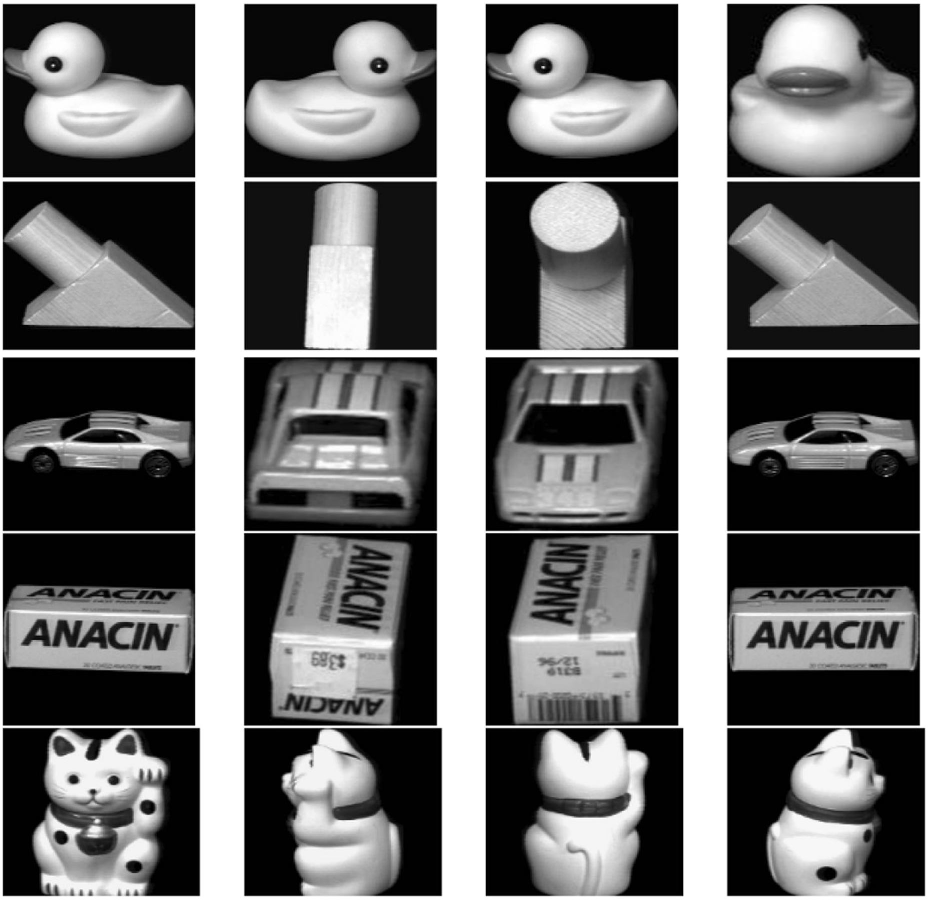



**Fig. 9** Sample prototype used in COIL-20 database

affects the recognition of the objects. The training set is prepared by taking the number of equally spaced views for each object. The remaining objects are taken as the testing images [5]. Figure 9 shows the proto types used for the coil data bases.

The prototype images were taken as the reference image instead of single reference image. The input image is compared with all the reference images and the minimal error is taken for the classification. The nearest neighbor classification is used for classification. The misclassified images are shown in the Fig. 10. The 47 images were misclassified out of 1200 images. The error rate is 3.91%.

## 4 Conclusion and future work

The proposed method produces a better result with reduced timing. The complexity of the program is very less. The proposed method gives good performance with occulted and noisy

**Fig. 10** The misclassified COIL data base images

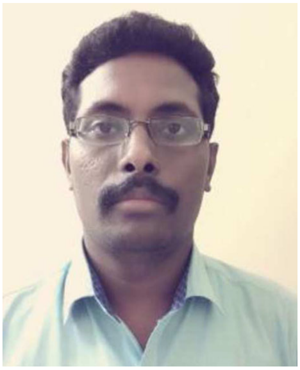
images. It can also be extended to the real time images, 3-D images, and industrial database. It can also extend to recognize multiple objects on the motion pictures. The proposed method has the wide application areas like military areas, investigation departments and industrial applications. In the future work some optimization methods are used to find the distance between the points in the recognition. The current work doesn't easily apply to real time applications which are left as future work. The computational complexity of the system is considered as the scope of the future.

## References

1. Ali SS, UsmanGhani M (2014) Handwritten digit recognition using DCT and HMMs. 12th International Conference on Frontiers of Information Technology, vol. 1, p 303–306
2. Amit Y, Geman D, Wilder K (1997) Joint induction of shape features and tree classifiers. IEEE Trans Pattern Anal Mach Intell 19(11):1300–1305
3. Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61(3):183
4. Bao P, Zhang L, Wu X (2005) Canny edge detection enhancement by scale multiplication. IEEE Trans Pattern Anal Mach Intell 27(9):1485–1490
5. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(4):509–522
6. Calabi E, Olver P, Shakiban C, Tannenbaum A, Haker S (1998) Differential and numerically invariant signature curves applied to object recognition. Int J Comput Vis 26:107–135
7. Coimbra MT, Cunha JS (2006) MPEG-7 visual descriptors—contributions for automated feature extraction in capsule endoscopy. IEEE Trans Circuits Syst Video Technol 16(5):628–637
8. Forssén PE, Lowe DG (2007) Shape descriptors for maximally stable extreme regions. In 2007 IEEE 11th International Conference on Computer Vision, p 1–8

9. Forsyth D, Mundy J, Zisserman A, Brown C (1991) Projectively invariant representations using implicit algebraic curves. Image Vis Comput 9(2):130–136
10. Gool LV, Moons T, Pauwels E, Oosterlinck A (1992) Semi-differential invariants. In: Mundy J, Zisserman A (eds) Geometric invariance in computer vision, p 193–214
11. Hong B-W, Soatto S (2014) Shape matching using multiscale integral invariants. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(1):25–44
12. Hung WL, Yang MS (2004) Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance. Pattern Recogn Lett 25(14):1603–1611
13. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
14. Li SZ (1999) Shape matching based on invariants. In: Progress in Neural Networks: Shape Recognition, vol. 6, p 203–228
15. McIlhagga W (2011) The canny edge detector revisited. Int J Comput Vis 91(3):251–261
16. Nadernejad E, Sharifzadeh S, Hassanpour H (2008) Edge detection techniques: evaluations and comparisons. Appl Math Sci 2(31):1507–1520
17. Prasad BG, Biswas KK, Gupta SK (2004) Region-based image retrieval using integrated color, shape, and location index. Comput Vis Image Underst 94(1–3):193–233
18. Ranzato M, Huang FJ, Boureau Y, LeCun Y (2007) Unsupervised learning of invariant feature hierarchies with applications to object recognition. Proc. of Computer Vision and Pattern Recognition Conference (CVPR 2007)
19. Sapiro G, Tannenbaum A (1993) Affine invariant scale space. Int J Comput Vis 11(1):25–44
20. Zhao C, Shi W, Deng Y (2005) A new Hausdorff distance for image matching. Pattern Recogn Lett 26(5): 581–586





**Mr. K. Senthil Kumar** received his Master's Degree in Electrical and Electronics Engineering from Anna University. He is a member of IAENG. His areas of interest include Digital Image Processing, Pattern Recognition, and Soft computing. He has published 5 papers in National and International Conferences and Journals. He has 9 years of experience in Teaching. He is a reviewer for many Journals and Conferences. He is currently working as Assistant Professor, Department of EEE, Tamilnadu College of Engineering, Coimbatore, Tamil Nadu, India.

**Dr. T. Manigandan** received his Master's Degree in Electrical Engineering and Doctor of Philosophy from Anna University, Chennai. He is a member of IEEE, ISTE, IE(I) and IETE. His areas of interest include Control System, Digital Image Processing, Pattern Recognition, Computer Vision, Soft computing. He has published more than 100 papers in National and International Journals. He has 28 years of experience in Teaching. He has Produced 14 Ph.Ds and currently 5 research scholars pursuing Ph.D. He is a reviewer for many Journals and Conferences. He is currently working as Principal and Professor, P. A. College of Engineering and Technology, Pollachi, Tamil Nadu, India.

**Dr. D. Chitra** received her Master's Degree in Computer Science and Engineering and Doctor of Philosophy from Anna University, Chennai. She is a member of IEEE, ISTE, CSI, IAENG and IRED. Her areas of interest include Digital Image Processing, Pattern Recognition, Computer Vision, Data Mining and Grid and Cloud Computing. She has published 62 papers in National and International Conferences and Journals. She has 15 years of experience in Teaching. She has guided 67 projects in both UG and PG, and currently 6 research scholars pursuing Ph.D. She is a reviewer for many Journals and Conferences. She is currently working as Professor and HOD in the Department of Computer Science and Engineering, P. A. College of Engineering and Technology, Pollachi, Tamil Nadu, India.

**Dr. L. Murali** received his Master's Degree in Electronics and Communication Engineering and Doctor of Philosophy from Anna University, Chennai. He is a member of ISTE, IE(I) and IETE. His areas of interest include VLSI Design, Digital Image Processing, Pattern Recognition, and Soft computing. He has published more than 15 papers in National and International Journals. He has 9 years of experience in Teaching. He is a reviewer for many Journals and Conferences. He is currently working as Associate Professor, Department of ECE, P. A. College of Engineering and Technology, Pollachi, Tamil Nadu, India.