Static Hand Gesture Detection and Classification Using Contour Based Fourier Descriptor

[1]Mrs.Anjali R.Patil [2]Dr. Mrs. S. Subbaraman
[1]Assistant Professor, Electronics Department [2]Professor, Electronics Department
[1]DKTE's Textile and Engineering Institute [2] Walchand College of Engineering
[1]Ichalkaranji [2] Sangli.
Maharashtra

Abstract: **Hand gesture recognition is a vital task as its application ranges from human-machine interface to sign language recognition. This paper present the empirical results of static hand gesture detection based on well known technique used for shape detection using Fourier Descriptors (FDs). Fourier descriptors are important features for the classification of two-dimensional connected shapes. Further the classification is done using multiclass Support Vector Machine (SVM). The performance is evaluated on Sebstian Marcel Database and qualified to provide average classification accuracy of 93%.**

Keywords: **Static hand gesture, Fourier Descriptors, Support Vector Machine, Classification Accuracy.**

I. INTRODUCTION

Hand gesture recognition plays an important role in wide area covering the applications from virtual reality to sign language recognition. The use of gestures in human computer interaction (HCI) has been extensively developed [1, 2]. Static gestures and dynamic gestures are the broad classes of hand gesture. Constant hand shapes are Static gestures and dynamic gestures are usually represented according to movements of hand. Dynamic gestures are supposed to change over period of time, while a static gesture is constant over time. Waving hand to say good bye is a simple example of dynamic hand gesture and the stop sign is an example of static gesture. To extract the complete information from the gestures, it is necessary to understand all the static and dynamic gestures over a period of time. Gesture recognition means recognizing and interpreting frames of dynamic gestures from the input data. Since dynamic gestures are variation of static gestures over time, recognizing static gestures is one of the most important parts in gesture recognition. Extracting powerful descriptors to characterize static gestures is a critical process.

Extensive research work is carried out by the researchers to extract features of static gestures. Shape features are extensively used for classification of static hand gestures in the literature. Human understand any scene which is composition of different object by their shapes. Apart from this classification of different segments in an image can be done based on the shape. For any classification generally images are first segmented into individual objects then they are exploited to perform object recognition. Whether it is classification of chromosomes of cell in biomedical field or may be recognition of character on each page of the books

to assist the blind people, object recognition is used. Once the object is segmented from the image, the next stage is to recognize or classify that object to appropriate class. If the main information for description or classification can be found in the boundary of the object it is natural to keep only the boundary for further analysis of the object. The situations where the shape features can be used are, in the classification of silhouettes of airplanes, photograph taken from the ground, classification of silhouettes of satellites, character recognition and in machine part recognition and many more.

The object is generally described by its boundary in a meaningful manner. Since each boundary is composition of collection of all connected curves, the concentration is upon the description of connected curves. Number of techniques are available to describe such curves, Fourier Descriptors are widely used important set of features [3] [4]. Other methods include chain encoding proposed by H. Freeman [5] and the polygonal approximation used by T. Pavlidis [6]. They determined the boundary as a connection of closed curves or segments. Contour-based shape descriptors include Fourier Descriptor (FD) [3, 4], wavelet descriptors [7], curvature scale space [8] and shape signatures [9]. Generally it is found that Contour based shape features are used to classify the different shaped object, but it is rarely used to describe the internal properties i.e. interior of the objects.

In hand gesture recognition, the techniques which provides unique features that are used primarily for shape representation as well as its time complexity is less is chosen so that the recognition of static hand gestures can be done in real time. It is also expected that the technique used should be invariant to translation, rotation, and scaling.

Autoregressive modeling and Fourier descriptors of closed contours was used in shape classifications work carried out by Hannu Kauppinen et.al. in [10]. The robustness of the system was tested using two sets of input data sets: images of letters and airplanes. The average accuracy reported was 99.5% using different Fourier descriptors.

In [11] Fourier descriptors were computed using complex contour technique by Persoon et.al. This experiment was carried out for character recognition and machine part recognition. The recognition accuracy was not reported by the authors.

Comparison between Fourier descriptors (FDs) and short-time Fourier descriptors (SFDs) was presented by Dengsheng Zhang et.al. in [12]. It has been found that FDs outperforms SFDs in terms of retrieval performance. The precision rate for FDs reported was 90% and that of SFDs was 85%.

In [13], S. Bourennane et.al. experimented different shape features for an online video application of vision-based hand posture recognition. The features were extracted using Fourier Descriptor, Hu and Zernike moments. The authors first used a benchmark database. They have created their own gesture vocabulary. They acquired 11 gestures from 18 persons and around 1,000 images per gesture and per person to evaluate the robustness of the features with respect to the scale changes, translation, rotation and viewpoint invariance. Recognition rate with Triesch database and Euclidean Distance observed was 78.76%.

In [14], Chin-Chen Chang et.al presented a new approach for recognizing static gestures. They used Zernike moments (ZMs) and Pseudo-Zernike moments (PZMs). The hand boundary is first extracted and a Minimum Bounding Circle (MBC) is calculated over the silhouette. The hand silhouette segmented into the finger and the palm by morphological operations as per the radius of the MBC circle. The best recognition rate observed was 97.3%; the worst recognition rate was 95.0%.

In image retrieval also shape features/descriptors plays an important role. Most of existing shape descriptors were usually either application dependent or non-robust. This makes them undesirable to be used for generic shape description. In [15], Dengsheng Zhang et.al. Presented a Generic Fourier Descriptor (GFD) to overcome the limitation of existing shape classification techniques. The method explains the shape descriptor which was derived by performing 2-D Fourier transform on a polar image. The acquired shape descriptor was independent of application as well as robust to translation and scale changes. Experimental results show that the proposed GFD outperforms common contour-based and region-based shape descriptors.

In this paper a method is proposed for classification of static hand gesture using Fourier Descriptors and the Support Vector Machines. The rest of this paper is organized as follows: Section 2 presents discussion on Fourier Descriptors followed by brief theory on Support Vector Machine in Section 3. Experimental set up and results are presented in Section 4. Conclusions are provided in Section 5.

II. FOURIER DESCRIPTORS

2-D Fourier transformation is extensively used for shape representation and analysis. The coefficients calculated by applying Fourier transform on the input image forms the Fourier descriptors of the shape. These descriptors generally represent the shape in a frequency domain. The global features of the shape are given by the low frequency descriptors and finer details of the shape are given by the higher frequency descriptors. The number of coefficients obtained after transformation are generally large, some of them are sufficient to properly define the overall features of the shape. High frequency descriptors that are generally used to provide the finer details of the shapes are not used for discrimination of the shape, so they can be ignored. By doing this, the dimensions of the Fourier descriptors used for capturing shapes are significantly reduced and the size of feature vector is also reduced.

As shape is connected object and is described using a closed contour that can be represented as a collection of the pixel coordinates in x and y direction. The coordinates can be considered to be sampling values. Suppose that the boundary of a particular shape has P pixels numbered from 0 to P - 1. The p-th pixel along boundary of the contour has position (x_p, y_p). The contour can be described using two parametric equations:

$$\mathbf{x(p)= x_p}$$
$$\mathbf{y(p)= y_p} \qquad \qquad \textbf{(1)}$$

The Cartesian coordinates of the boundary pixel is not considered as Cartesian coordinates instead they are converted to the complex plane by using the following equation

$$\mathbf{s(p)= x(p)}+ i\mathbf{y(p)} \qquad \textbf{(2)}$$

The above equation means that the x-axis is treated as real axis and y-axis as imaginary axis of a sequence of complex numbers. Although the interpretation of the sequence was recast, the nature of the boundary itself was not changed. Of course this representation has one great advantage: It reduces a 2-D to 1-D problem. The Discrete Fourier Transform of this function is taken and frequency spectra are obtained. Discrete Fourier transform of s(p) is given by

$$a(u) = \frac{1}{P} \sum_{k=0}^{p-1} s(p) e^{-j2\pi up/P} \qquad (3)$$

Where u= 0, 1,2, ….P-1.

The complex coefficients a(u) are called the Fourier descriptors of the boundary. The inverse Fourier transform of these coefficients restores s(p) and given by the following equation:

$$S(P) = \frac{1}{P} \sum_{u=0}^{p-1} a(u) e^{-j2\pi p/P} \qquad (4)$$

Where p=0, 1, 2, ….P-1

Formation of Feature vector using Fourier Descriptor is done using following steps:
1. Fourier descriptors Input static hand gesture color image.
2. Preprocess the input image using following steps:
   i) Convert RGB image to YUV color space.
   ii) Extract the skin color by applying thresholding.
   iii) Noise removal by applying median filter.
   iv)Convert the color image to black and white image with gesture representing binary as 1 and background as 0.
3. Noise (eg. Spikes) is removed using morphological filters.
4. Crop, and resize the extracted hand gesture to a size of 40 x 40 pixels in order to have the dimension of the Fourier descriptor uniform for all the images in the training database.
5. Extract the boundary pixel coordinates of the gesture image.
6. Represent the boundary in the complex plane where the column-coordinate is the real part and the row-coordinate the imaginary part.
7. Calculate Fourier transform and form the feature vector of the 20 dimensional

The invariance to the translation is obtained by - nullifying the 0-th Fourier descriptor (position invariance), dividing all Fourier descriptors by the magnitude of the 1-st Fourier descriptor (size invariance) and only considering the magnitude of the Fourier descriptors (orientation and starting point invariance). By applying this normalization, the 0-th and 1-th Fourier descriptors do not provide any information. Hence these are eliminated. Experimentally we have chosen number of coefficients of Fourier descriptors to

be 20. These descriptors, invariant to scale, translation and rotation, form the feature vector

III. SUPPORT VECTOR MACHINE

Support Vector Machine (SVM), proposed in 1990's by Vapnik is based on the linear discriminate functions used for classification and regression [16]. The fundamental concept of the SVMs is to build a hyper plane as the decision making plane, which is used to separate the binary classes with positive (+1) and negative (-1) with the largest margin, which is computed by finding the minimum Vapnik Chervonenkis (VC) dimension of SVM. In case of the binary classification problem the feature extraction is required to be initially performed. In the proposed method the 20 dimensional Fourier Descriptors are used for training SVM. Let us take the notations for the training data $xj \in \Re^d$ with a label $yj = \{-1,+1\}$, for all the training data j = 1, 2,3, .………… l , where l is the number of data, and d is the dimension of the classification problem. When the two classes are linearly separable in $\Re^d$, separating hyper plane which gives the smallest generalization error among the infinite number of possible hyper-planes is required to be computed. Such an optimal hyper-plane gives the maximum margin between the two classes. The margin is calculated as the sum of the distances from the hyper-plane to the closest data points of each of the two classes. These closest data points are called Support Vectors (SVs). The middle line on Fig. 1 represents the optimal separating hyper-plane. Hyper-plane is represented with a direction vector w (normal to the hyper-plane) and an offset vector b that satisfies the equation
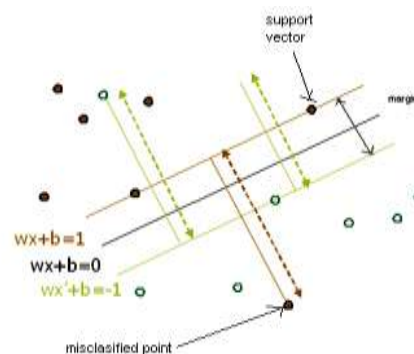
$$w^t . x + b = 0 \qquad (5)$$

Fig.1. Linear separating hyper-plane and Support vectors. The separation problem is to determine the hyper-plane such that $x_i + b \geq 1$ for positive examples and $wx_i + b \leq 1$ for negative examples.

In case of binary classification in the two-dimensional case, there are support lines, instead of planes, and the decision boundary also is a line as shown in the figure. The distance from any point x on the line $w_x + b = 1$ to the decision boundary is given by

$$\frac{f(x)}{\|x\|} = \frac{1}{\|w\|} \qquad (6)$$

In the above, note that f(x) = 1 for any x on the line $wx_i + b = 1$. Similarly, the distance from a point on the line $wx_i + b = -1$ to the decision boundary is given by $\frac{-1}{\|w\|}$. So, the distance between the two supporting lines, or the margin, is $\frac{2}{\|w\|}$. Margin is maximized by maximizing $\frac{2}{\|w\|}$ or minimizing $\frac{\|w\|}{2}$ or for the sake of simplicity as in calculus, $\frac{\|w\|^2}{2}$

$$\min_{w,b} \Phi(w) = \frac{\|w\|^2}{2} \qquad (7)$$

The optimal hyper-plane is found by minimizing (6) under the constraint (7) to correctly classify the training data.

$$y_i = (w.x_i + b) - 1 \geq 0, \forall i \quad (8)$$

Implementing linear SVM described above is a Quadratic Programming (QP) problem for which standard techniques like Lagrange Multipliers can be used.

$$f_\lambda : \Re^n \rightarrow \{-1, +1\} \qquad (9)$$

The set of functions f_λ could be a set of Radial Basis Functions or a multi-layer neural network.

IV. EXPERIMENTAL SET-UP AND RESULTS

For performing the static hand gesture classification shape based features are calculated as mentioned in the section II and the classification accuracy is calculated. The comparison is done based on various performance parameters like Classification Rate (CR) or Classification Accuracy (CA), False Alarm Rate (FAR), False Rejection Rate (FRR).

In training database total 6 postures A, B, C, Point, Five and V from Sebastian Marcel Database [17] have been used. Sample postures are given in Fig.2:

Fig.2. Six postures 'a', 'b', 'c', 'five', 'point', 'v' from Sebastian Marcel static hand postures database[17].

300 samples were used in the training phase (50 samples of each posture) and in testing 600 samples were used (100 samples of each posture). Out of 100 samples in the testing, 50 were used for calculation of False Acceptance Rate (FAR) and remaining 50 were used for finding False Rejection rate (FRR). The size of posture images selected was 100 x 100. Fig. 3 below shows the result of the algorithm which was used to extract the features of posture 'A', using the steps given in section II. Fig. 4 shows the boundary detection of the different hand postures mentioned above.
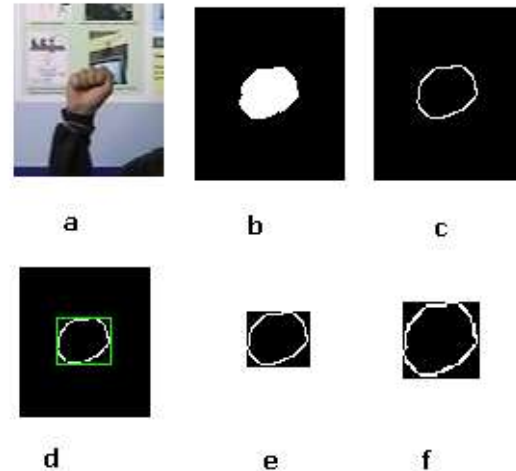

Fig.3. Steps in Feature extraction of sample posture 'A' a).input RGB image b)skin detection c) edge detection using Canny operator d) Bounding Box to the hand shape e)Hand cropped image e) Resized hand image of size 40x40
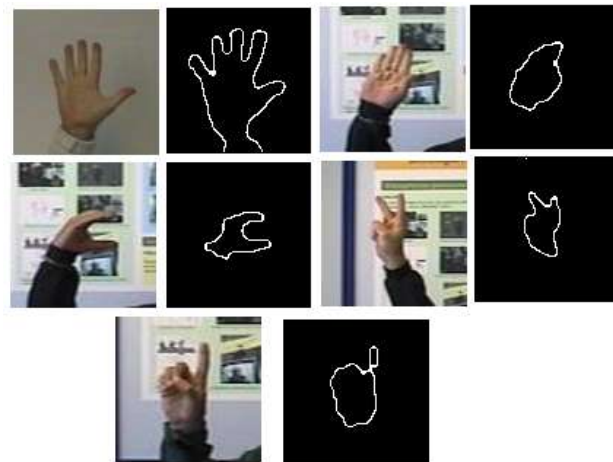
Fig.4.Boundary Detection results of the proposed system for posture Five, B, C, V, Point.

In proposed system, the classifier used to classify six static hand gestures (and not only two) was the combination of binary SVMs in a Classifier Tournament structure. For every pair of classes, a binary classifier was trained that classifies between them. The features of the hand postures are generated using the algorithm mentioned in section II. The features were entered into all the binary classifiers. Every binary classifier outputs a class number giving a specific hand posture. Voting is taken and maximum number was the final class of posture. Detailed discussion on different multiclass SVM can be found at [18].

A. *Performance parameters*:

1) False acceptance rate (FAR): is a measure of probability how a classification system will accept an incorrect input as a positive match. A system's FAR typically is stated as the ratio of the number of false acceptances divided by the number of identification attempts.

$$FAR(\%) = \frac{FalseMatches}{impostorAttempts} \qquad (10)$$

2) False Rejection rate (FRR): By contrast the FRR is a measure of probability a system will incorrectly reject an input as a negative match. A system's FRR typically is stated as the ratio of the number of false rejections divided by the number of identification attempts.

$$FRR(\%) = \frac{FalseNonMatch}{EnroleeAttempt} \qquad (11)$$

3) Classification accuracy: It is defined as the rate of positive decisions in the total number of decisions for each stage of the procedure for instance positive match for all the test image will lead to 100% Recognition Accuracy. The action or process of classifying the correct postures from the test database is Recognition rate.

$$CA(\%) = \frac{PositiveMatch}{EnroleeAttempt} \qquad (12)$$

Table 1 shows the confusion matrix obtained from SVM. Total 100 images of each posture are used for testing the machine. Table 2 represents the FAR, FRR and CA for the six postures.

TABLE I.

ABLE 1. CONFUSION MATRIX (SVM AS CLASSIFIER)

Postures	A	B	C	POINT	FIVE	V
A	**90**	-	6	-	-	4
B	1	93	-	6	-	-
C	-	-	93	7	-	
POINT	-	-	2	91	-	7
FIVE	1	-	3	-	92	4
V	3	3	-	-	-	94

TABLE II.

ABLE 2. CLASSIFICATION ACCURACY

Classifier	Static Hand Gesture	No. of test images	FAR %	FRR %	CA %
SVM	A	50	5	10	90
	B	50	3	7	93
	C	50	11	7	93
	Point	50	13	9	91
	V	50	15	8	92
	Five	50	0	6	94
Average Classification Accuracy					92.16

V. CONCLUSION AND FUTURE SCOPE

The experiential results using SVM classifiers for static hand gesture classification for the Fourier descriptors are reported in this paper. And its performance is given with reference to CA, FAR and FRR. 20 dimensional feature vectors are formed using Fourier Descriptors which are obtained by applying Fourier transform on boundary pixels of the posture. These features are invariant to translation, rotation, scaling and change of initial point.
The average classification accuracy observed is 92.16%.

Database that has been used in this paper consist of the posture images having different view and illumination with complex background. The classification accuracy can be further increased by selecting the robust features which will be invariant to view and illumination.

REFERENCES

[1]. Vladimir I. et.al. 'Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review', IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, July 1997.

[2]. Sushmita Mitra, Tinku Acharya, 'Gesture Recognition: A Survey' IEEE Transactions on Systems, Man, and

Cybernetics—Part C: Applications and Reviews, Vol. 37, No. 3, MAY 2007.

[3]. D. S. Zhang and G. Lu Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study. In Proc. of IEEE International Conference on Multimedia and Expo (ICME2001), August 22-25, 2001, Tokyo, Japan, pp.317-320.

[4]. Charles T. Zahn and Ralph Z. Roskies. Fourier Descriptors for Plane closed Curves. IEEE Trans. On Computer, c-21(3):269-281, 1972.

[5]. H. Freeman, A. Saghri, Generalized chain codes for planar curves, in: Proceedings of the Fourth International Joint Conference on Pattern Recognition, Kyoto, Japan, November 7–10, 1978, pp. 701–703

[6]. T. Pavlidis, Algorithms for shape analysis of contours and waveforms, IEEE Trans. Pattern Anal. Mach. Intell. 2 (4) (1980) 301–312.

[7]. Q. M. Tieng and W. W. Boles. Recognition of 2D Object Contours Using the Wavelet Transform Zero-Crossing Representation. IEEE Trans. on PAMI 19(8): 910-916 Aug.1997

[8]. Farzin Mokhtarian and Alan Mackworth. Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes. IEEE PAMI-8(1):34-43, 1986

[9]. E. R. Davies. Machine Vision: Theory, Algorithms, Practicalities. Academic Press, 1997.

[10]. Hannu Kauppinen, Tapio Seppanen and Matti Pietikainen. An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification. IEEE Trans. PAMI-17(2):201-207, 1995

[11]. Eric Persoon and King-sun Fu. "Shape Discrimination Using Fourier Descriptors". IEEE Trans. On Systems, Man and Cybernetics, Vol.SMC-7(3):170-179, 1977.

[12]. Dengsheng Zhang and Guojun Lu "A comparison of shape retrieval using fourier descriptors And short-time fourier descriptors" Gippsland School of Computing and Information Technology Monash University Churchill, Victoria 3842

[13]. S. Bourennane · C. Fossati "Comparison of shape descriptors for hand posture recognition in video" Received: 25 July 2009 / Revised: 29 July 2010 / Accepted: 30 July 2010 / Published online: 15 August 2010 © Springer-Verlag London Limited 2010

[14]. Chin-Chen Chang, Jiann-Jone Chen1, Wen-Kai Tai2 And Chin-Chuan Han "New Approach for Static Gesture Recognition" Journal Of Information Science And Engineering 22, 1047-1057 (2006)

[15]. Dengsheng Zhang and Guojun Lu "Shape Based Image Retrieval Using Generic Fourier Descriptors"

[16]. Takefumi Matsunaga and Oshita Masaki, "Recognition of Walking Motion Using Support Vector Machine", Proc. of ISICE 2007, pp. 337-342, 2007.

[17]. S. Marcel, "Hand posture recognition in a body-face centered space," in Proc. Conf. Human Factors Comput. Syst. (CHI), 1999, pp. 302Processing, 2005, vol. 2, pp. 954 – 957

[18]. Kai-Bo Duan, S. Sathiya Keerthi, 'Which Is the Best Multiclass SVM Method? An Empirical Study', Springer Berlin, 2005.