Automatic Indian Sign Language Recognition System

Karishma Dixit
Department of Computer Engineering and Applications
GLA University
Mathura, India
karishma.dixit@gmail.com

Anand Singh Jalal
Department of Computer Engineering and Applications
GLA University
Mathura, India
anandsinghjalal@gmail.com

Abstract— **Sign Language is the most natural and expressive way for the hearing impaired. This paper presents a methodology which recognizes the Indian Sign Language (ISL) and translates into a normal text. The methodology consists of three stages, namely a training phase, a testing phase and a recognition phase. Combinational parameters of Hu invariant moment and structural shape descriptors are created to form a new feature vector to recognize sign. A multi-class Support Vector Machine (MSVM) is used for training and recognizing signs of ISL. The effectiveness of the proposed method is validated on a dataset having 720 images. Experimental results demonstrate that the proposed system can successfully recognize hand gesture with 96% recognition rate.**

Keywords— Indian Sign Language (ISL), Multi-class Support Vector Machine (MSVM)

I. INTRODUCTION

Sign Language is the means of communication among the deaf and mute community. Sign Language emerges and evolves naturally within hearing impaired community. Sign Language communication involves manual and non-manual signals where manual signs involve fingers, hands, arms and non-manual signs involve face, head, eyes and body. Sign Language is a well-structured language with a phonology, morphology, syntax and grammar. Sign language is a complete natural language that uses different ways of expression for communication in everyday life. Sign Language differs from other languages as it has no spoken word. The structure of the spoken language makes use of words sequentially whereas a sign language makes use of numerous body movements in parallel. Sign Language recognition system transfers the communication from human-human to human-computer interaction. Sign language interpreters are used by deaf and dumb people to communicate with the hearing world. The aim of the sign language recognition system is to present an efficient and accurate mechanism to transcribe text or speech, thus the "dialog communication" between the deaf and hearing person will be smooth. There is no standardized sign language for all deaf people across the world. However, sign languages are not universal, as with spoken languages, these differ from region to region.

There are two main approaches used in the sign language recognition that is Glove/Device based and Vision based. In the glove based method the user has to wear a device which carries a load of cables so as to connect the device to a computer. Such devices are expensive and reduce the naturalness of the sign language communication. In contrast, the Vision based method requires only a camera and directly deals with image gestures. It is a two step process: sign capturing and sign analysis. Vision based methods provide a natural environment to the user and reduces the complications as in the glove based method.

Every country has its own sign language with a high level of grammatical variations. The sign language exists in India is commonly known as Indian Sign Language (ISL). It has been argued that perhaps the same sign language is used in Nepal, Sri Lanka, Bangladesh, and border regions of Pakistan [1]. Examples of other sign languages are the American Sign Language (ASL), the British Sign Language (BSL), the Korean Sign Language (KSL), and so on. In general, the semantic meaning of the language components in all sign languages differs, however there are signs with a universal syntax. For instance, a simple gesture by one hand expressing "hi" or "goodbye" has the similar meaning across the world and in all forms of sign languages. ISL is a complete natural language, found in India with its own morphology, phonology, syntax, and grammar [1]. Indian Sign language (ISL) is a visual-spatial language providing linguistic information through hand, arms, face, and head/body gestures. ISL produces both isolated as well as continuous signs. An isolated sign focuses on a single hand gesture, and is an exacting hand configuration and pose represented by a particular image. A continuous sign is a moving gesture, represented by series of images.

The objective of the proposed work is to efficiently recognize the signs of Indian Sign Language (ISL) and translate the accurate meaning of those recognized signs. In this paper, we have proposed an Indian sign language recognition method, which is based on combination of invariant moment and shape descriptor features. We are considering only manual signs which comprises of hand gesture of isolated signs. Results show that combined approach provides an effective recognition rate with a high accuracy.

The paper is organized as follows: In section II proposed methodology is explained. In section III the results are discussed and finally, section IV concludes the paper.

978-1-4673-4529-3/12/$31.00 ©2012 IEEE

II. PROPOSED METHODOLOGY

Fig. 1 shows the framework of the proposed system. It is clear from Figure 1 that the proposed system consists of three phase's via. a training phase, a testing phase and a recognition phase. In the training phase, each class is trained with a multi-class support vector machine (MSVM). Hu invariant moment and structural shape descriptors are combined to make a combinational feature vector that are to be extracted from the input image in the testing phase after applying preprocessing. In the recognition phase, different classes are used for testing an input gesture. The outcome with the most probable group is identified to recognize the gesture. Finally, after the recognition of input image their meaning is displayed on the screen.

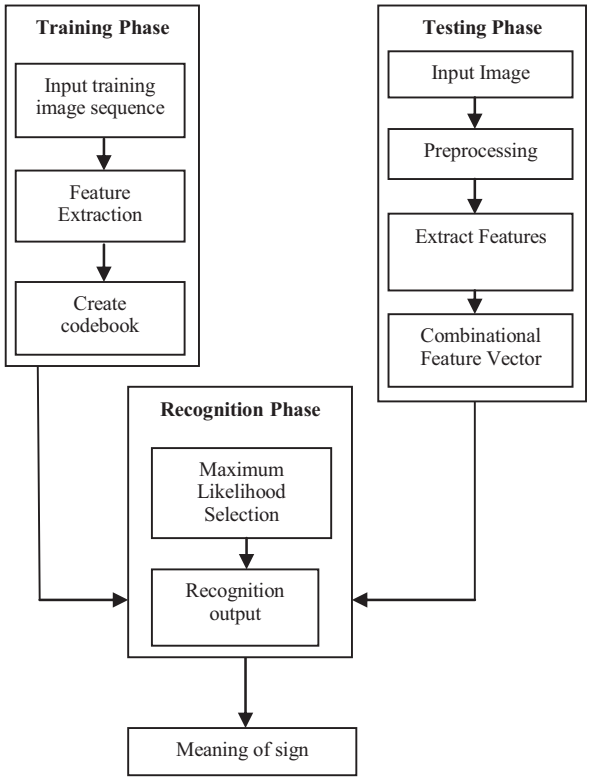


Fig. 1. Block diagram of hand gesture recognition system

A. Preprocessing

Preprocessing is applied to images before extracting features from hand images. Preprocessing consists of two steps, segmentation and filtering. All the functions are applied on a gray scale image. The segmentation of an input image of a hand gesture is performed using Global thresholding algorithm [2]. Global thresholding algorithm tackles any segmentation problem as classification problem, and the image level is divided into two classes one is hand and the other is background. The assigned value for hand pixel is "1" and the value for background pixel is "0".

After converting a gray scale image into binary image we have to make sure that there is no noise in image so we use filtering technique. To eliminate noise from the segmented image a median filter is applied on the image. The segmentation and filtering is applied below:

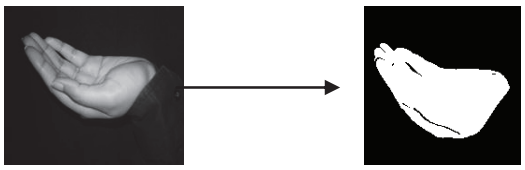


Fig. 2. Segmentation of gray scale image

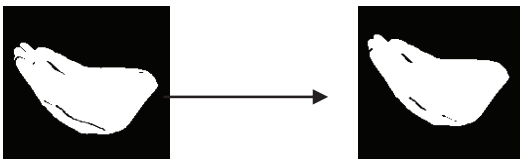


Fig. 3. Filtering of Segmented image

B. Feature Extraction through Hu Invariant Moment

In [3] the author proposed a moment known as, Hu invariant moment" which is derived from the theory of algebraic invariant. Hu invariant moment is used for scale and position invariant pattern identification. The advantage of using Hu invariant moment is that it can be used for disjoint shapes. In particular, Hu invariant moment set consists of seven values computed by normalizing central moments through order three [3]. In terms of central moment the seven moments are given as below:

$$M_1 = \eta_{20} + \eta_{02} \tag{1}$$

$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{2}$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{3}$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{4}$$

$$\begin{aligned} M_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{5}$$

$$\begin{aligned} M_6 = {} & (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \\ & + \left[4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})\right] \end{aligned} \tag{6}$$

$$\begin{aligned} M_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{30})^2(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \tag{7}$$

where the normalized central moment η_{pq} ,

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{pq}^{\gamma}} \quad , \quad \gamma = \left(\frac{p+q}{2}\right)+1$$

And $$\mu_{pq} = \sum_{x}\sum_{y}\left(x-\bar{x}\right)^{p}.\left(y-\bar{y}\right)^{q}.f\left(x,y\right)$$

where, x = 0,…..., M-1 ; y = 0,……, M-1 ;

p, q = 0,1,2,3,….

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad ; \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

$m_{oo} = area\ of\ subject$

$m_{0,1}, m_{1,0} = centre\ of\ mass$

These seven values given by Hu are used as feature vector for each hand gesture.

C. Feature Extraction through Structural Shape Descriptors

The shape descriptors used in the proposed approach are as follows:

- **Aspect ratio:** It is the ratio of major axis to minor axis [4]. The value of this ratio is always between 0 and 1. It is a measurable property of a boundary and is sensitive to the elongation of a boundary. It is as follows:

$$aspect\ ratio = \left(\frac{\text{major axis}}{\text{minor axis}}\right) \tag{8}$$

- **Compactness:** It shows the image similarity to its circumference with respect to its centre. Its maximum value is 1 for circles and its value decreases for elliptical shapes [4]. It is defined as:

$$Compactness = \frac{4\Pi.Area_{image}}{Perimeter_{image}^{2}} \tag{9}$$

- **Solidity:** It is a measure of ruggedness of a boundary profile. The solidity value 1 indicates a region with no concavities in its boundary and a value less than 1 indicates a region with indentations, or containing holes [4]. It is defined as:

$$Solidity = \frac{Area_{image}}{Area_{convex\ hull}} \tag{10}$$

- **Elongation:** It is a measure of height and width of a rotated minimal bounding box. Its maximum value is 1 for a circle and is less for other shapes [5]. It is defined as:

$$Elongation = \frac{4.Area}{\Pi.(major\ axis)^{2}} \tag{11}$$

- **Spreadness**: It indicates that how the image is spread over the background [5]. It is defined as:

$$Spreadness = \frac{\mu_{20}+\mu_{02}}{\mu_{00} * \mu_{00}} \tag{12}$$

- **Orientation:** It is a measure of overall direction of the image shape [6].

D. Multi-class Support Vector Machine

In the proposed work, a multi-class support vector machine has been used to appropriately classify the hand gestures among multiple classes. In the proposed approach a binary classifier is converted into multiclass classifier. The ij^{th} binary classifier uses the pattern of class i as positive and the patterns of class j as negative examples [7]. For the final outcome, minimum distance of the generated vector has been calculated to the binary pattern representing each class. Suppose there are three classes such as, *x, y, z*. Train the binary classifier differentiating two classes at a time, such as $x \times y$, $x \times z$, $y \times z$. Each class receives the unique identifier as shown in Table 1. First, perform the binary comparison $x \times y$ and tag the class x with outcome +1, class y with outcome -1 and set the remaining entries in that column to 0. The entry 0 represents a "Don't care" value. Then repeat this procedure for the remaining classes.

TABLE 1. UNIQUE IDS OF CLASSES

	$x \times y$	$x \times z$	$y \times z$
x	+1	+1	0
y	-1	0	+1
z	0	-1	-1

After completing the above process, we get a vector, which further used to calculate the minimum distance from this vector to each one of the class unique IDs.

E. Recognition

The images in dataset are trained through MSVM. When an input image is given in the testing phase first the preprocessing is applied to the input image. Then the hu invariant moment and structural shape descriptor features of an input image are calculated and store in a codebook. After that, the features of the input image are matched with the codebook through MSVM and the most likelihood image is recognized and retrieved with their meaning.

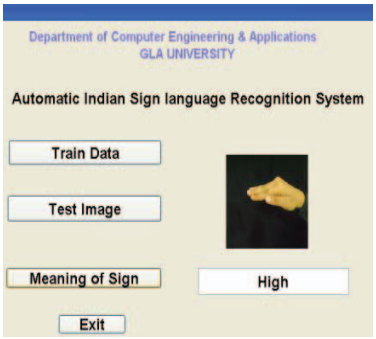
III. RESULT AND DISCUSSIONS

For proving the effectiveness and accuracy of the proposed system, we have carried out a number of experiments. A sign image dataset has been created, which contains 720 images. The dataset is classified as a testing set that has 600 images and a training set that has 120 images. 720 images include different

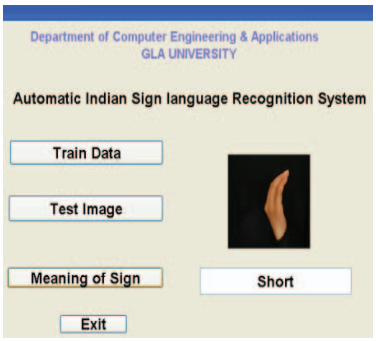
signs from ISL of 12 different people. These are divided into 60 classes. The accuracy of the proposed system is evaluated as follows [8]:

$$Accuracy = \left(\frac{correctly\ classified\ gestures}{total\ no.\ of\ gestures} \right) \times 100\% \quad (13)$$

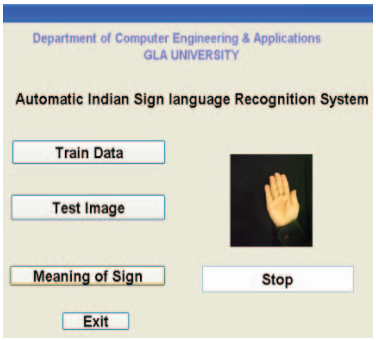
The resolution of grabbed image is 256*256. Each image is resized by a magnification factor of 0.5 to reduce the computation process. All operations are performed on gray scale image.



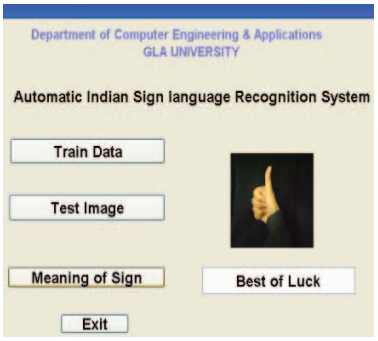
(a)



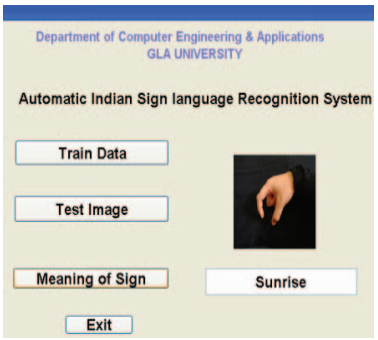
(b)



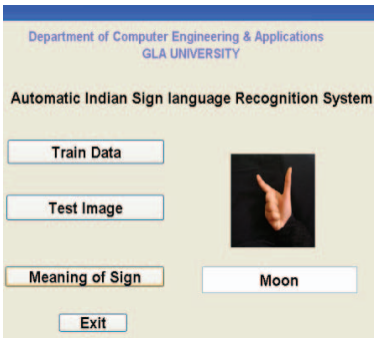
(c)



(d)



(e)



(f)

Fig. 4. Snapshots of the proposed system

Fig. 4 shows the graphical user interface (GUI) of the proposed system. For training the gesture images, the train data button is pressed and the system is trained through multiclass support vector machine. After that, an input image should be given by clicking the test image button and finally the meaning of sign is retrieved by clicking that button. In (a) an input image whose original meaning is „high" should be given to the system for testing and from the result window we can see that the image is accurately identified and its meaning is retrieved by the system. The actual and retrieved meaning of this image is identical. Similarly, the images in (b)-(f) are correctly identified and their actual meanings are displayed on the result window. From (a)-(f), all the images are recognized and their actual meaning are displayed on the result window with a good accuracy.

A number of experiments have been conducted with 60 different sign and we have found that the success rate of our

approach in classification reaches up to almost 96.23%. Table 2 illustrates the experiments carried with the feature vector with one or more group of features. It is evident from Table 2 that the proposed approach outperforms than other methods.

TABLE 2. RECOGNITION ACCURACY

Experiment No.	Feature Group	% correctly classified instances
1	Invariant moments	40.36%
2	Structural Shape Descriptors	80.98%
3	Proposed	96.23%

IV. CONCLUSION

In this paper, an automatic ISL recognition system has been created which works on real-time basis and employs the use of combinational feature vector with a MSVM classifier. In the combinational feature vector, Hu invariant moment and structural shape descriptors are used collectively for achieving better recognition results. The use of MSVM increases recognition performance. Results demonstrate that the combination of invariant moments and shape descriptors gives better result, as shape descriptors define the boundary of an image while the invariant moments are invariant to change in scale and position of an image. The future directions will focus on signer independent, large vocabulary systems in both isolated and continuous recognition tasks.

REFERENCES

[1] T. Dasgupta, S. Shukla, S. Kumar, S. Diwakar and A. Basu, "A Multilingual Multimedia Indian Sign Language Dictionary Tool," The 6‟ th Workshop on Asian Language Resources, 2008, pp. 57-64.

[2] N. Otsu, "A Threshold Selection method from Gray level Histograms," IEEE Transactions on Systems, Man, and Cybernetics., vol. 9, no. 1, pp. 62–66, 1979.

[3] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," IRE Transactions on Information Theory, vol. 8, pp. 179-187, 1962.

[4] N. Jamil, Z. A. Bakar and T. M. T. Sembok, "Image Retrieval of Songket Motifs using Simple Shape Descriptors," IEEE Proceedings of the Geometric Modeling and Imaging-New Trends, 2006, pp. 171-176.

[5] T. R. Trigo and M. R. S. Pellegrino, "An Analysis of Features for Hand Gesture Classification," In 17th International Conference on Systems, Signals and Image Processing, 2010, pp. 412-415.

[6] M. Panwar and P. S. Mehra, "Hand Gesture Recognition for Human Computer Interaction," In International Conference on Imgae Information Processing, 2011, pp. 1-7.

[7] A. Rocha, D. C. Hauagge, J. Wainer and S. Goldestine, "Automatic Fruit and Vegetables Classification from Images." Computer and Electronics in Agriculture, vol. 70, pp. 96-104, 2010.

[8] S. Meena, A study on Hand Gesture Recognition Techniques, Project report, 2011.