A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic

P. V. V. Kishore and P. Rajesh Kumar

Abstract—This paper proposes a complete skeleton of isolated Video Based Indian Sign Language Recognition System (INSLR) that integrates various image processing techniques and computational intelligence techniques in order to deal with sentence recognition. The system is developed to improve communication between hearing impaired people and normal people promising them better social prospects. A wavelet based video segmentation technique is proposed which detects shapes of various hand signs and head movement in video based setup. Shape features of hand gestures are extracted using elliptical Fourier descriptions which to the highest degree reduces the feature vectors for an image. Principle component analysis (PCA) still minimizes the feature vector for a particular gesture video and the features are not affected by scaling or rotation of gestures within a video which makes the system more flexible. Features generated using these techniques makes the feature vector unique for a particular gesture. Recognition of gestures from the extracted features is done using a Sugeno type fuzzy inference system which uses linear output membership functions. Finally the INSLR system employs an audio system to play the recognized gestures along with text output. The system is tested using a data set of 80 words and sentences by 10 different signers. The experimental results show that our system has a recognition rate of 96%.

Index Terms—Indian sign language, fuzzy inference system, wavelet transform, canny edge operator, image fusion, elliptical fourier descriptors, principle component analysis.

I. INTRODUCTION

The sign language is natural language used for communication by hearing impaired people. A sign language relates letters, words, and sentences of a spoken language to hand signs and human body gestures facilitating hearing impaired people to communicate among themselves. Sign language recognition systems provide a channel for communication between hearing impaired people and normal people. By making this system fully realizable can create jobs for hearing impaired people in different areas of their interest. Advances in sign language recognition can largely promote research in the areas of human computer interface. This paper provides a novel technique to recognize signs of Indian sign language using wavelet transform and fuzzy inference system.

The principal constituent of any sign language recognition system is hand gestures and shapes normally used by deaf people to communicate among themselves. A gesture is defined as a energetic movement of hands and creating signs with them such as alphabets, numbers, words and sentences. Gestures are classified into two type static gesture and dynamic gestures. Static gesture refer to certain pattern of hand and finger orientation where as dynamic gestures involve different movement and orientation of hands and face expressions largely used to recognize continuous stream of sentences. Our method of gesture recognition is a vision based technique which does not use motion sensor gloves or colored gloves for the system to recognize hand shapes. A complete gesture recognition system requires understanding of hand shapes, finger orientations, hand tracking and face expressions tracking.

Accordingly sign language recognition systems are classified in to two broad categories: sensor glove based [1], [2] and vision based systems [3]. The first category requires signers to wear a sensor glove or a colored glove. The wearing of the glove simplifies the task of segmentation during processing. Glove based methods suffer from drawbacks such as the signer has to wear the sensor hardware along with the glove during the operation of the system. In comparison, vision based systems use image processing algorithms to detect and track hand signs as well as facial expressions of the signer, which is easier to the signer without wearing gloves. However, there are accuracy problems related to image processing algorithms which are a dynamic research area.

Thad starner proposed a real time American Sign Language recognition system using wearable computer based video [4] which uses hidden makov model (HMM) for recognizing continuous American Sign Language system. Signs are modeled with four states of HMMs which have good recognition accuracies. Their system works well but it is not signer independent. M.K.Bhuyan [5] used hand shapes and hand trajectories to recognize static and dynamic hand signs from Indian sign language. The used the concept of object based video abstraction technique for segmenting the frames into video object planes where hand is considered as a video object. Their experimental results show that their system can classify and recognize static, dynamic gestures along with sentences with superior consistency.

Yu Zhou, Xilin chen [6] proposed a signer adaptation method which combines maximum a posteriori and iterative vector field smoothing to reduce the amount of data and they have achieved good recognition rates.

In this paper we are proposing a sign language recognition system for transforming signs of Indian sign language in to voice commands using hand and head gestures of humans.

Discrete wavelet transform (DWT) based fusion algorithm [7], [8] is used to extract edges from the video stream of different signers. Further hand and head shape features are extracted from these binary videos using Elliptical Fourier descriptors [9]-[11]. Fourier descriptor data is further processed using Principle component analysis (PCA) to reduce the size of the feature vector. We have considered 80 different gestures of Indian sign language. Finally the reduced feature vector is used to train the fuzzy inference system [12]-[17]. Testing of the system is done by considering samples from 10 different signers for each gesture. The recognized signs are connected to audio output of the computer using MATLAB software along with text output. Finally a GUI is built to train and test new signs automatically by users.

II. DATABASE REPRESENTATION OF INDIAN SIGN LANGUAGE

Indian sign language does not have a standard database like American Sign Language database (ASL) that can accessed from the internet. Hence we have determined to collect our Indian sign language (ISL) database by collaborating closely with Indian deaf society and shanthi ashram school for deaf children, Visakhapatnam. In the process initially we have collected 80 signs of alphabets, numbers and words from Indian sign language. The list of some of the words is shown in tableI.

TABLE I: SOME INDIAN SIGN LANGUAGE GESTURES ALONG WITH THEIR ENGLISH MEANINGS.

S No.	Indian Sign Language Video Frame	English Meaning
1.		COW
2.		TEN
3.		WITHOUT

| 4. | | SPARROW |
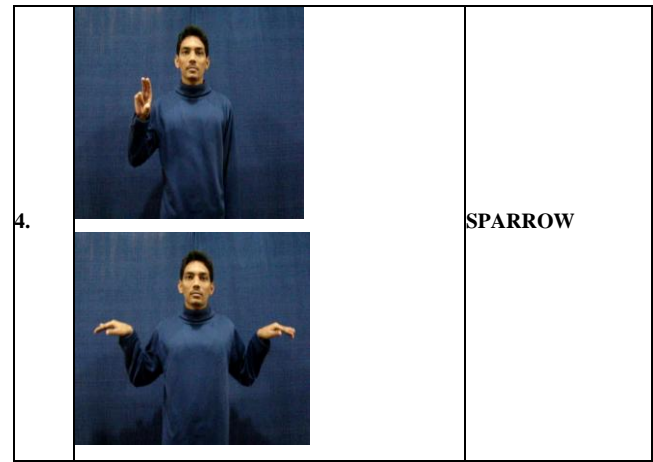
The videos are shot in a controlled environment using a Sony cyber shot H7 digital camcorder having a resolution of 640X480 pixels capturing 25 frames per second. For our experiment ten different signers volunteered where each signer is asked to repeat each gesture two times under different lighting conditions with a total number of 1600 gesture videos for a total of 80 signs. A video acquisition process is subjected to many environmental conditions such as position of the camera, illumination and noise. The only constraint of our system is that the signer must wear a dark colored full sleeves shirt with a dark background.

III. PROPOSED SYSTEM DESIGN

The proposed system has four processing stages explicitly video preprocessing, image segmentation, feature extraction and pattern classification. Fig. 1 shows the framework of the proposed method.

A. Stage 1: Video Preprocessing

Preprocessing stage consists of basically two operations namely image resizing and filtering. RGB video is converted into indexed image frames. Indexed Video frames are resized to reduce in resolution to lower the processing time. Here we scaled down the resolution from 640×480 to 256×256 image
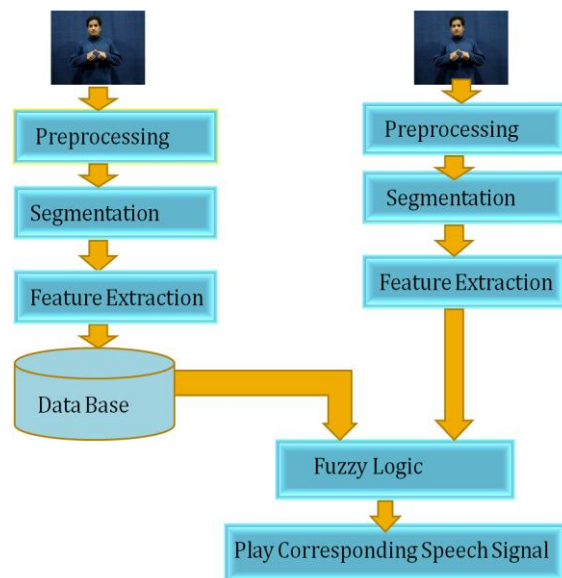

Fig. 1. Proposed framework of our system.

using interpolation methods. RGB images consist of 3 planes

and processing involves manipulating all three planes simultaneously which consumes processing time. Hence RGB indexed video frames are converted to gray scale indexed video frames by forming a weighted sum of the R, G and B components to reduce the complexity.

Further Gaussian low pass filter is applied to the gray scale image to remove the high frequency noise and smoothen the image frames. Fig. 2 shows all the preprocessing techniques applied to indexed frames of sign 'A'.

Fig. 2. Resizing, RGB-to-Gray and Gaussian filtering operations in preprocessing stage.

B. Stage 2: Segmentation

Segmentation on non trivial video frames decides the success or failure of any computerized analysis procedures. Segmentation is done by fusing discrete wavelet transform (DWT) and canny operated images. This fusion technique gives unambiguous true edges compared to any edge detection algorithm alone.

On the filtered image apply canny edge detector. Canny edge detector performs the following operations on the image like, image smoothing with a Gaussian filter, computing gradient and angle, applying non maxima suppression to the gradient of the image, use double thresholding and connectivity analysis to detect and link edges. The canny edge detected image applied to single frame of signer is shown in Fig. 3. As we can observe some unwanted lines near hands and neck which are false edges concerning our experiment. We are interested in extracting head and hand portions for further processing which is achieved by fusing canny with discrete wavelet transform.
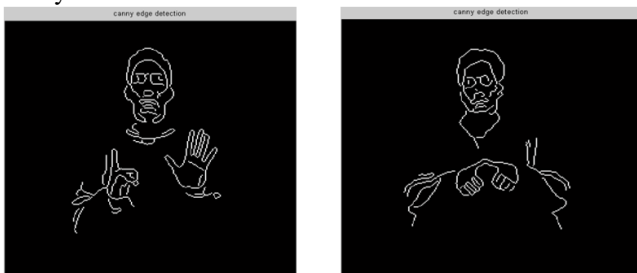

Fig. 3. Canny operated image frames of signs TEN and alphabet 'A' of two different signers.

Adopting image decomposition model of wavelet transform, the original gray scale image can be subdivided into low frequency, high frequency and diagonal information. Further the low frequency sub image is sub divided into low frequency, high frequency and diagonal images using daubechies family of wavelets. The real edges of the image are embedded in high frequency components of the image along with noise. Fuse Low frequency sub image reconstructed wavelet coefficients and canny image by summing. On the obtained fused image frame thresholding is applied using Otsu's algorithm resulting in an edged image.

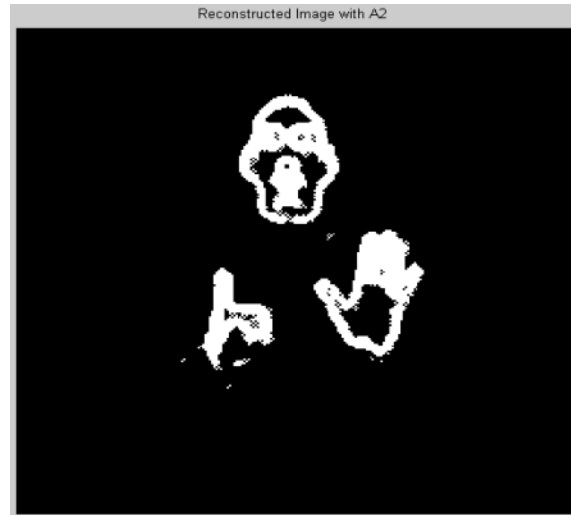The result of the segmentation is shown in Fig. 4.

Fig. 4. Result of Segmentation obtained by fusion of canny and DWT.

C. Stage 3: Feature Extraction

The segmentation yields a low data in the form of pixels along a boundary or pixels contained in a region. We are mainly interested in the boundary pixels that are hand shapes of the signer. To extract shape outline with minimum number of pixels for an image frame without losing shape information we choose Fourier Descriptors. Fourier descriptors often attribute to early work by cosgriff (1960), allows us to bring the power of Fourier theory to shape description. The basic idea of Fourier descriptors is to characterize a curve by a set of numbers that represent the frequency content of a whole shape. The Fourier descriptors allow us to select a small set of numbers that describe a shape for an image frame. This property Fourier descriptors is helpful because these Fourier coefficients carry shape information which is not insensitive to translation, rotation and scale changes. But the changes in these parameters can be related to transformations on descriptors.

Fourier representations are articulated in terms of orthogonal basis functions, causing the parameters to be distinct and hence avoid redundancy. The boundary of hands and head or its contour can be represented by closed curves. As a result of periodicity of curves projection on the vertical and horizontal axis we used elliptical fourier representations to model hand and head contours. The elliptical fourier descriptors of a closed contour is defined as

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ c_0 \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \begin{bmatrix} \cos(kt) \\ \sin(kt) \end{bmatrix} \quad (1)$$

where

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} x(t)dt \qquad c_0 = \frac{1}{2\pi} \int_0^{2\pi} y(t)dt$$

$$a_k = \frac{1}{2\pi} \int_0^{2\pi} x(t)\cos(kt)dt \quad b_k = \frac{1}{2\pi} \int_0^{2\pi} x(t)\sin(kt)dt$$

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} y(t)\cos(kt)dt \quad d_k = \frac{1}{2\pi} \int_0^{2\pi} y(t)\sin(kt)dt$$

Thus any closed curve can be modeled by its fourier coefficients (a_0 , c_0 , a_k , b_k , c_k , d_k). Where k represents the rank of the ellipse and for $k=1$ corresponds to fundamental component of the closed curve. To fit exactly on the image closed curve one has to employ more than one ellipses to do so.

From geometric point of view a simple ellipse is modeled by the equation

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \times \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} \qquad (2)$$

where A is semi major axis oriented in horizontal axis direction and B is semi minor axis oriented in vertical axis. The model starting point lies with semi major axis. Considering rational angle and phase shift from major axis we obtain a more comprehensive representation of the ellipse as

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = E \times \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} \qquad (3)$$

where

$$E = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \times \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \times \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$$

θ = Rotational Angle

φ = Phase shift of the ellipse respectively

This geometric interpretation of ellipse gives us better visualization. For K^{th} ellipse as shown in the Fig. 5.
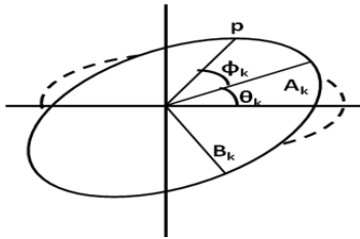

Fig. 5. K^{th} Ellipse showing parameters.

Therefore Eq(3) becomes

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \times \begin{bmatrix} \cos(kt) \\ \sin(kt) \end{bmatrix} \qquad (4)$$

$$E_k = \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \qquad (5)$$

From the above equation we can write the fourier coefficients of kth (a_k, b_k, c_k, d_k) as

$$a_k = A_k \cos\theta_k \cos\phi_k - B_k \sin\theta_k \sin\phi_k$$

$$b_k = -A_k \cos\theta_k \sin\phi_k - B_k \sin\theta_k \cos\phi_k$$

$$c_k = A_k \sin\theta_k \cos\phi_k + B_k \cos\theta_k \sin\phi_k$$

$$d_k = -A_k \sin\theta_k \sin\phi_k + B_k \cos\theta_k \cos\phi_k$$

where A_k, B_k, θ_k, φ_k are more understandable parameters of the same ellipse and the relationship between these set parameters could be computed. The plot of different fourier coefficients along with the view of head and hand contours is shown in Fig. 6.
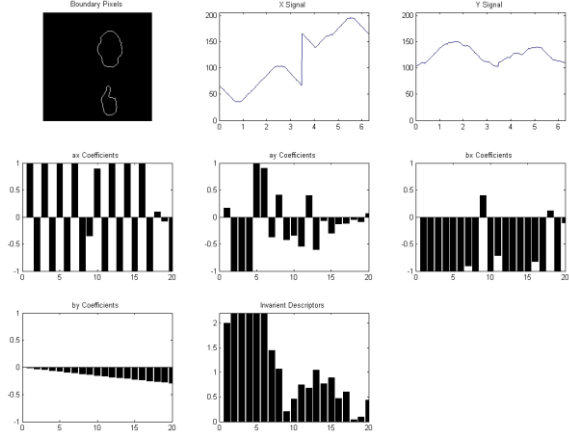

Fig. 6. Plot shows boundary pixels of a signers head and hand shapes considering 20 coefficients and different Fourier descriptors.

For a typical video of the signer with 45 frames we obtain a matrix of 45×20 values representing the shape information of hands and head. For 80 signs, our feature vector size becomes very large and consumes more processing time in the next stages. Hence we treat the current feature vector using principle component analysis (PCA) which uses factorization to transform data according to its statistical properties. This data transformation is particularly useful for classification problems when data set is very large. In the obtained feature matrix each column of the matrix defines a feature vector. For classification problems we select the features with large values of variance.

To begin with, the database is a 80×80 matrix consisting of 80 feature vectors and each column representing a particular sign video. This feature matrix will be used to train fuzzy inference system.

D. Stage 4: Pattern Classifiacation

For pattern classification we considered Takagi-sugeno-kang (TSK) or simply sugeno type Fuuzy inference system [18]-[23] because the output membership functions are linear or constsnt. Sugeno fuzzy inference system consists of fiv steps, fuzzification of input variables, applying fuzzy 'and' or 'or' operator, calculating the rule weights, calculating the output level and finally defuzzification. Many methods are proposed to generate fuzzy rule base. The basic idea is to study and generate the optimum rules needed to control the input without compromising the quality of control. In this paper the generation of fuzzy rule base is done by subtractive clustering technique in sugeno fuzzy method for classification video. Cluster center found in the training data are points in the feature space whose neighborhood maps into the given class. Each cluster center is translated into a fuzzy

rule for identifying the class. A generalized type-I TSK model can be described by fuzzy IF-THEN rules which represent input output relations of a system. For multi input single output first order type-I TSK can be expressed as

IF xi is Q_{1k} and x_2 is Q_{2k} and ... and x_n is Q_{nk}, THEN Z is

$$W = p_0 + p_1\chi_1 + p_2\chi_2 + \dots + p_n\chi_n \qquad (6)$$

where x_1, x_2…. x_n and Z are linguistic variables; Q_{1k}, Q_{2k} …and Q_{nk} are the fuzzy sets on universe of discourses U.,V. and W, and $p_0, p_1, p_2, \dots, p_n$ are regression parameters.

With subtractive clustering method, X_j is the jth input feature x_j ($j \in [1, n]$), and Q_{jk} is the MF in the kth rule associated with j^{th} input feature. The MF Q_{jk} can be obtained as

$$Q_{jk} = \exp\left[-\frac{1}{2}\left(\frac{X_j - \chi_{jk}}{\sigma}\right)^2\right] \qquad (7)$$

where x_{jk} is the $\sigma = \sqrt{\frac{1}{2a}}$ j^{th} input feature of x_k, the standard deviation σ of Gaussian MF given as

$$\text{Re}\,cognition\,Rate(\%) = \frac{Number\,of\,Correctly\,Classified\,Signs}{Total\,Number\,of\,Signs} \times 100 \qquad (8)$$

Table II summarizes different gestures used for the analysis of our proposed system.

Classification of different gestures is done using Fuzzy inference system for 80 different gestures of Indian sign language by 10 different signers. Table III gives details of the fuzzy inference system used for gesture classification. Fig. 7 shows the input membership functions. The performance of the system is evaluated based on its ability to correctly classify signs to their corresponding speech class. The recognition rate is defined as the ratio of the number of correctly classified signs to the total number of signs:

TABLE II: SUMMARY OF SIGN LANGUAGE USED.

All English Alphabets, Numbers, cow, donkey, duck, peacock, fat, feather, foolish, funny, nest, what, crow, young, upwards, towards, come, good, love, mother, father, where are you going, do your home work etc.

TABLE III: DETAILS OF FUZZY INFERENCE SYSTEM USED FOR GESTURE CLASSIFICATION.

Name	'fis_inslr'
Type	sugeno'
And Method	'min'
Or Method	'max'
Defuzz method	'wtaver'
Imp Method	'prod'
Agg Method	'sum'
Input	[1x80 struct]
Output	[1x1 struct]
Rule	[1x25 struct]

Table IV shows the results obtained when training 10 samples for each gesture with different signers. Table shows recognition rates of some signs used for classification.
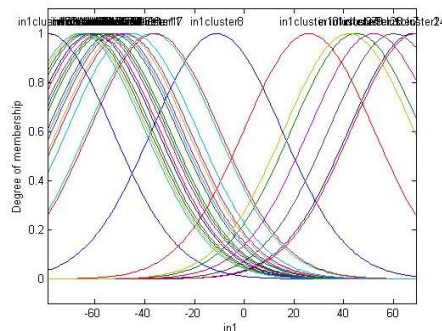

Fig. 7. Input membership functions.

The total number of signs used for testing is 80 from 10 different signers and the system recognition rate is close to 96%. The system was implemented with MATLAB version 7.0.1.

IV. CONCLUSIONS

In this paper we developed a system for recognizing a subset of the Indian sign language. The work was accomplished by training a fuzzy inference system by using features obtained using DWT and Elliptical fourier descriptors by 10 different signer videos for 80 signs with a recognition rate of 96%. In future we are looking at developing a system for Indian sign language that works in real-time.

TABLE IV: RESULTS OBTAINED WHEN TRAINING SIGNS WITH 10 DIFFERENT SIGNERS.

Sign	Correctly Recognized Signs	False Recognition	Recognition Rate (%)
A	10	0	100
B	10	0	100
C	10	0	100
D	10	0	100
X	10	0	100
M	8	2	80
N	9	1	90
Y	7	3	70
Cow	9	1	90
Duck	10	0	100
Crow	6	4	60
Fat	9	1	90
Feather	8	2	80
Love	9	1	90
Together	10	0	100
Come here	7	3	70
Do your Home Work	6	4	60
Numbers 1-10	100	0	100
Upwards	10	0	100
Total	258	22	92.142

REFERENCES

[1] G. Fang and W. Gao, "Large vocabulary contineous sign languagre recognition based on trnsition-movement models," *IEEE Transaction on Systems,MAN, and Cybernetics*, vol. 37, no. 1, pp. 1-9, January 2007.

[2] G. Grimes, "Digital data entry glove interface device, patent4,414,537," *AT &T Bell Labs*,1983.

[3] T. Starner and A. Pentland, "Real-time american sign language recognition from video using hidden markov models," *Technical Report, MIT Media Laboratory Perceptual Computing Section*, Technical Report number. vol. 375, 1995.

[4] M.-H. Yang and N. Ahuja, " Extraction of 2D motion trajectories and its aapplication to hand gesture recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1061-1074, August 2002.

[5] M. K. Bhuyan and P. K. Bora, "A frame work of hand gesture recognition with applications to sigb language," *Annual India Conference, IEEE*, pp.1-6.

[6] Y. Zhou and X. L. Chen, "Adaptive sign language recognition with Exemplar extraction and MAP/IVFS," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 297-300, March 2010.

[7] G. Pajares, "A wavelet-based image fusion tutorial," *Pattern Recognition*, vol. 37, no. 10, pp. 1855-1872, 2004.

[8] J. N. Ellinas and M. S. Sangriotis, "Stereo image compression using wavelet coefficients morphology," *Image and Vision Computing*, vol. 22, no. 2, pp. 281-290, 2004.

[9] F. P. Kuhl and C. R. Giardina, "Elliptic fourier descriptors of a closed contour," *CVGIP*, vol. 18, pp. 236–258, 1982

[10] C. C. Linand R. Chellappa, "Classification of partial 2D shapes using fourier descriptors," *IEEE Trans. PAMI*, vol. 9, no. 5, pp. 686–690, 1987.

[11] E. Persoon and K.-S. Fu, "Shape description using fourier descriptors," *IEEE Trans. SMC*, vol. 3, pp. 170–179, 1977

[12] M. Sugeno, "An introductory survey of fuzzy control," *Inform. Sci.*, vol. 36, pp. 59-83, 1985.

[13] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller-part-I and part-II," *IEEE Trans. Syst., Man, Cybem.*, vol. 20, no. 2, pp. 404-435, 1990.

[14] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. 15, no. 1, pp. 116-132, 1985.

[15] L. X. Wang and I. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cyben.*, vol. 22, no. 6, pp. 1414-1427, 1992.

[16] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Trans. Fuzzy Systems*, vol. 1, no. 1, pp. 7-31, 1993.

[17] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets and syst.*, vol. 52, pp. 21-32, 1992.

P. V. V. Kishore (SMIEEE'07) received his M.Tech degree in electronics from Cochin University of science and technology in the year 2003, and currently pursuing PhD at Andhra University College of engineering in Department of ECE from 2008. He is working as research scholar at the Andhra university ECE department. He received B.Tech degree in electronics and communications engineering from JNTU, Hyd. in 2000. His research interests are digital signal and image processing, computational intelligence, human computer interaction, human object interactions. He is currently a student member of IEEE.

Dr. P. Rajesh Kumar (MIEEE'09, FIETE'02) received his Ph.D degree from Andhra University College of Engineering for his thesis on Radar Signal Processing in 2007. He is currently working as associate professor at Dept. of ECE, Andhra University College of engineering, Visakhapatnam, Andhra Pradesh. He is also Assistant Principal of Andhra University college of Engineering, Visakhapatnam, Andhra Pradesh. He as produced numerous research papers in national and international journals and conferences. He has guided various research projects. His research interests are digital signal and image processing, computational intelligence, human computer interaction, radar signal processing.