

Video Audio Interface for Recognizing Gestures of Indian Sign Language

P.V.V.Kishore

*Research Scholar/ECE
Andhra University College of Engineering
Visakhapatnam, 530003, INDIA*

pvvkishore@gmail.com

P.Rajesh Kumar

*Associate Professor/ECE
Andhra University College of Engineering
Visakhapatnam, 530003, INDIA*

rajeshauce@gmail.com

E.Kiran Kumar

*Postgraduate Student/ECE
Dadi Institute of Engineering and Technology
Visakhapatnam, 530017, INDIA*

kiraneepuri@gmail.com

S.R.C.Kishore

*Post Graduate Student/ECE
Pydah College of Engineering
Visakhapatnam, 530037, INDIA*

ravickishore@gmail.com

Abstract

We proposed a system to automatically recognize gestures of sign language from a video stream of the signer. The developed system converts words and sentences of Indian sign language into voice and text in English. We have used the power of image processing techniques and artificial intelligence techniques to achieve the objective. To accomplish the task we used powerful image processing techniques such as frame differencing based tracking, edge detection, wavelet transform, image fusion techniques to segment shapes in our videos. It also uses Elliptical Fourier descriptors for shape feature extraction and principal component analysis for feature set optimization and reduction. Database of extracted features are compared with input video of the signer using a trained fuzzy inference system. The proposed system converts gestures into a text and voice message with 91 percent accuracy. The training and testing of the system is done using gestures from Indian Sign Language (INSL). Around 80 gestures from 10 different signers are used. The entire system was developed in a user friendly environment by creating a graphical user interface in MATLAB. The system is robust and can be trained for new gestures using GUI.

Keywords: Sign language, Edge Detection, Wavelet Transform, Image Fusion, Elliptical Fourier Descriptors, Principle Component Analysis, Fuzzy Inference System.

1. INTRODUCTION

Sign language is seen as the core distinctive feature that defines a deaf community. The role of sign language recognition systems in the society is to ensure that deaf people have equality of opportunity and full participation in society. Sign language is represented basically by continuously varying different hand shapes and movements by a signer. Sign language recognition is decoding and understanding the information embedded in the hand shapes and converting them to meaning full words. Sign language is a natural language of the deaf community. By developing sign language recognition system a hearing impaired person can easily interact with a normal person at different levels in the society. We have attempted to design a basic system of sign language pertaining to Indian Sign Language.

Going back to early days of sign language Stokoe et.al [1] showed that signs are made of basic articulatory units initially referred to as cheremes commonly referred in these times as phonemes in resemblance with that to words in spoken language.

Logically sign language understanding consists of linguistic analysis of hands tracking, hands shapes, hands orientations, sign verbalization and also with important linguistic information communicated with head movements and facial expressions. Sign language is in many ways different from spoken language such as facial and hand terminology, references in virtual signing space, and grammatical differences as explained.

The major difficulty in sign language recognition compared to speech recognition is to recognize simultaneously [2] different communication attributes of a signer such as hands and head movement, facial expressions and body pose. All these attributes has to be considered simultaneously for a good recognition system.

The second major problem faced by sign language recognition system designers is tracking the signer in the clutter of other information available in the video. This is addressed by many researchers as signing space [3]. A sign language space can be created with entities such as humans or objects stored in it around a 3D body centered space of the signer. The entities are executed at a certain location and later referenced by pointing to the space. A major challenge faced by researchers to define a model for spatial information containing the entities created during the sign language dialogue.

Additional difficulties arise in the form of background in which signer is located. Most of the methods developed so far use simple backgrounds in controlled set-up such as simple backgrounds, special hardware like data gloves, restricted sets of actions, restricted number of signers, resulting different problems in sign language feature extraction.

For adapting the system for signer independent as in case of automatic speech recognitions systems which are robust being able to cope with different dialects. Speaker adaptation techniques known for speech recognition can be used to make the system more robust. While for recognizing signs of few signers only the intrapersonal variabilities in appearance and velocity of their hands needed to be modeled. As the number of signers increases, the quantity and diversity of variabilities is extremely increased.

In continuous speech recognition as well as continuous sign language recognition, co articulation effects have to be considered. Owing to location changes in virtual signing space, we have to encounter movement epenthesis problem [3,4]. Movement epenthesis refers to movements which occur regularly in natural sign language in order to change the location in signing space.

In automatic speech recognition the energy of the audio signal is used to detect silence in the sentences. New features and models have to be defined for silence detection in sign language recognition. The simplest way to do this is by motion analysis in the video with no reliability, because words are signed by just holding a particular posture in the signing space. In order to automate the sign language recognition system there should be a consistent silence detection and sentence boundaries adding to speed and recognition performance.

Sign language similar to spoken language is not confined to a particular region or territory. It is practiced differently around the world. In USA it is known as American Sign Language (ASL) [5, 6] and in Europe [7, 8], Africa [9] and in Asia [10, 11]. Unlike America and Europe sign language, India does not have a standard sign language with necessary variations, though. But recent past Ramakrishna missions Vivekananda University, Coimbatore came up with an Indian sign language dictionary. There are around 2037 signs [12] currently available in sign language.

Accordingly sign language recognition systems are classified in to two broad categories: sensor glove based [13] and vision based systems [14, 15]. 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 sterner proposed a real time American Sign Language recognition system using wearable computer based video [14] which uses hidden markov 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 [16] 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.

Rini Akmelia et.al [17] proposed a real time Malaysian sign language translation using color segmentation technique which has a recognition rate of 90%. Nariman Habili et.al [18] proposed a hand and face segmentation technique using color and motion cues for the content based representation of sign language video sequences. Yu Zhou, Xilin chen [19] 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.

Gesture recognition systems that are implemented statistical approaches [20], example based approaches [21], finite state transducers [22] have registered higher translation rate. In the last decade there were more efficient algorithms for training [23] and more efficient algorithms for generation [24] with the advancement of powerful computers and bigger parallel corpora.

The Essential Sign Language Information on Government Networks (eSIGN Project) [25] is a European project that has contributed in a large way in developing tools for automatic creation of sign language contents. The important breakthrough in this project has been 3D avatar (VGuido) to represent signs in the sign language resourcefully. The project also consists of visual settings for creating sign animations in rapid and easy way translating web content into sign language. This translation of web content in to sign language helps deaf people significantly in using the internet. Currently this project is translating websites in Germany, UK and the Netherlands.

Recently researchers are developing systems that can translate speech to sign language which require a parallel corpus to be able to train the language and translation models, and to evaluate the systems. The European Cultural Heritage Online Organization (ECHO) presents a corpus in British and Netherlands sign languages [26]. Another was made by The American Sign Language Linguistic Research Group at Boston University [5].

In this paper various image processing and artificial intelligence techniques are combined to generate a fully automated sign language recognition system. This system converts videos of signs by different signers into text and voice messages. The system can be trained to handle new signs at any point during operation. A graphical user interface (GUI) has been developed to make the system more user friendly. We named our system Indian Sign Language Recognition system (INSLR).



















The paper is organized as follows: in Section 2, creation of database representation of Indian Sign language. Section 3, presents an overview of the system along with different stages of design. Section 4, analysis of results and performance of the system under various conditions and finally section 5 summarizes the main conclusions.

2. DATABASE REPRESENTATIONS OF INDIAN SIGN LANGUAGE

In order to approach the problem of translating signs in to speech it is necessary to create a database of videos of different signs by multiple signers. Unlike American Sign Language or British sign language, Indian sign language does not have a standard database that is available for use. Hence we have created our own video database of Indian signs in collaboration with Indian Deaf Society [12] and shanthi ashram school for deaf children. Initially we have created 80 signs of alphabets, numbers, words and sentences by multiple signers of Indian sings.

The experimental setup consists of a dark background and the signer wearing a dark jacket with full sleeves. This controlled environment reduces tracking and segmentation problems. The RGB videos are acquired using a Sony cyber shot H7 digital cam coder having a resolution of 640×480 pixels capturing 25 frames per second. Higher resolution incorporates longer delays in the video acquisition process and higher execution times. The videos are acquired under normal lighting conditions to simulate real-time environment. Natural lightning of around 15-20 lux was present during the video acquisition process.

A total of nine different signers volunteered where each signer is asked to repeat the sign twice under different conditions with a total number of 1440 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. Some of the signs in our data base are presented in TABLE 1. The only constraint of our system is that the signer must wear a dark colored full sleeves jacket with a dark background.

Sl.No	Frame#10	Frame#20	Frame#30	English Representation of the Sign
1				I DON'T WANT TO GO
2				COW
3				UPWARDS
4				SMALL
5				CROW
6				DUCK

7				TEN
8				PEACOCK
9				FAT

TABLE 1: Shows different signs in our database displaying three Frames per video

3. DESIGN OF PROPOSED SYSTEM

The system design is based on four broad issues related as video preprocessing, image segmentation, feature extraction and pattern classification. Figure 1 shows the framework of our proposed sign language recognition system.

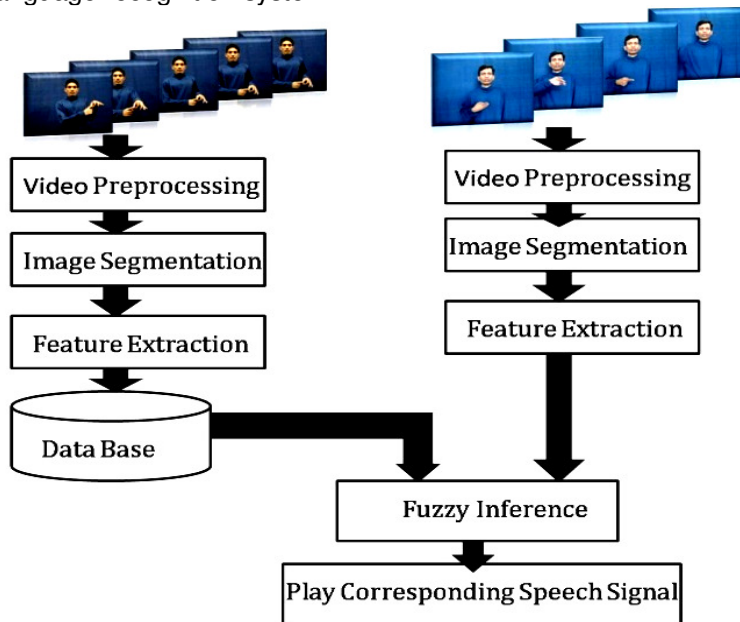


FIGURE 1: Framework of sign language recognition system

From the figure 1 we can understand the overall working process of the system. Video of signer is preprocessed using different techniques to avoid any false segmentation in the next stage. Image segmentation stage divides the entire frames of images of the signers into hands and head segments of the signer. Shape features are extracted and are optimized before saving to the database. This process is repeated for all the available signs using a video of a signer. The sign language database can be upgraded to add new signs into the database.

The system can now be tested with the signs present in the database using fuzzy inference system based on rules. The rule based system is a very powerful tool used by many researchers for pattern classification tasks. Once trained the fuzzy inference system produces a voice and text response corresponding to the sign irrespective of the signer and changing lighting conditions.

3.1 Video Preprocessing

Video preprocessing comprise of three operations on the input video sequence of the signer to make the segmentation process efficient. The first in the process is scaling the video form 640×480 pixels to 256×256 for reducing data and thereby saving the processing time. Scaling is done using bilinear interpolation algorithm, where the output pixel value is a weighted average of pixels in the nearest 2×2 neighbourhood.

The resized RGB (Red, Green, and Blue) video has 3 color planes per frame. Any minor alteration to the color video involves modification in all three R, G and B plane. Hence operating in RGB plane can decelerate the recognition process. This problem is solved by converting RGB color video into Grayscale video by eliminating the hue and saturation information and retaining the luminance. The grayscale image frame is formed by a weighted sum of R, G and B components.

During video acquisition the video is subjected to high frequency periodic noise naturally caused by the presence of electrical or electromechanical interference. Also moving objects in the video causes motion blur. These types of noise and blur can be eliminated by applying a Gaussian low pass filter. Convolving noisy blurred video frame with Gaussian filter point spread function gives a output filtered video frame. All the preprocessing steps are shown in figure2.

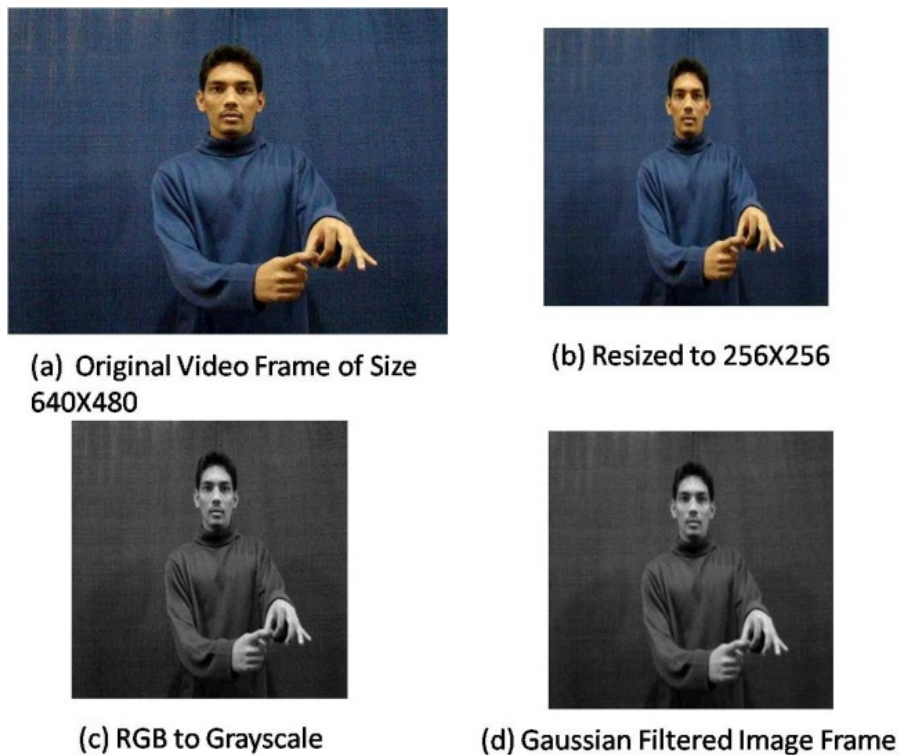


FIGURE 2: Video Preprocessing

3.2 Segmentation

One of the most serious challenges we faced is to automatically segment objects from a video sequence of the signer. Canny fused Wavelet based video object segmentation have turn out to be an indispensable part in achieving better segmentation, because it combines the necessary qualities of canny operator and two dimensional wavelet transform. Segmentation of hands and head portion is made perfect by applying a series of image processing techniques such as

morphology, canny operator, and wavelet transform and image fusion. These entire operations guarantee that hand shape in any orientation is perfect in each and every frame of the video.

A straightforward and proficient edge detection algorithm is proposed based on morphology and multilevel discrete wavelet transform (DWT). The edge detection process is a combination of gradient calculation, applying two dimensional discrete wavelet transform and finally calculating inverse wavelet transform for image reconstruction. Gradient image of canny edge detector and the approximations of wavelet transform are fused together. For the non-uniform illumination problem, a global thresholding algorithm (Otsu's Method), can be constructed to minimize the inter class variance of black and white pixels. By applying the algorithm on signer videos show that it not only detects edges correctly but also suppresses the noise to a great extent.

Edge is a result of gray level discontinuities. The edge detection of images is an important research area in the field of image processing. For long it has been an objective of the edge detection algorithms to emphasize problems to detect edges in the noisy environments, problem of false edge detection and problem of un-detection [27]. The most common approach for detection is by using 2D first and second order derivatives. The process is convolving 2D gradient with the image which returns a one for non-uniform regions and a zero for uniform regions. For high-quality segmentation, the first step is selection of suitable edge detection method.

Traditional edge detection techniques are based on the gradient operators, such as Sobel, Prewitt and Roberts. The most popular edge detectors such as laplacian of Gaussian and canny operators have the ability to detect true weak edges. Due to edge complexity, they are very sensitive to noise and brightness variations. Baris Sumengen, B. S. Manjunath [28] proposed a simple and efficient edge detection algorithm showing that using multiscale edge detection techniques eliminates the need for explicit scale selection and edge tracking. There are many different methods for edge detection; the traditional methods use first-order derivative operators to detect edges, such as Robert operator, Prewitt operator [29]. Wavelet has well local performance in time-frequency field and the characteristic of multi-resolution analysis, which is very suitable for image processing [30, 31].

Dilation and Erosion are two fundamental morphological operations. Dilation adds pixels to the borders of items in an image while erosion removes pixels on item borders. The number of pixels added or removed from the items in an image depends on the size and shape of the structuring element used to process the image. An essential part of the dilation and erosion operations is the structuring element used to probe the input image. A structuring element is a matrix consisting of only 0's and 1's that can have any arbitrary shape and size. The pixels with values of 1 define the neighbourhood. In this paper we have chosen a two line structuring element of length three which are oriented diagonally at +45 degrees and -45 degrees. The two structuring elements are shown in the figure 3.

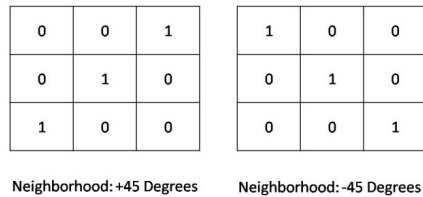


FIGURE 3: Flat Structuring elements containing 3 elements

In segmenting videos dilation and erosion are used in combination to produce a binary gradient image before applying discrete wavelet transform. Our algorithms dilate and erode the original image by using the structuring elements defined in the figure1 and subtracts the dilated image from the eroded image. The result of above operation applied to image frames is shown in figure 4.

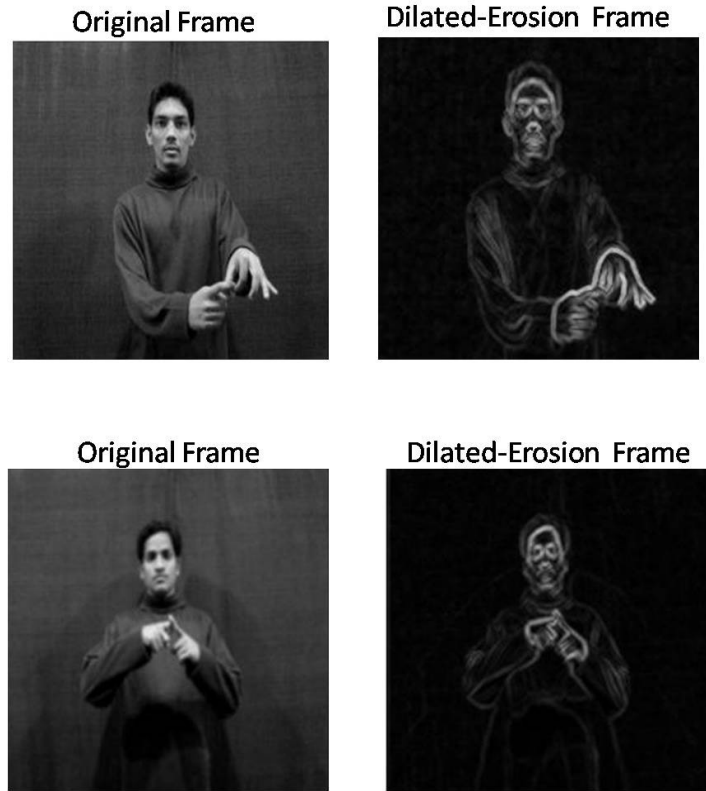


FIGURE 4: Result of dilation-erosion

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it. Wavelet transform is defined as decomposing a signal into a family of functions that are the translation and dilation of a unique function $\theta(x)$ called the basic wavelet. A function $\theta(x)$ is a basic wavelet if its average is zero. $L^2(\mathbb{R})$ denotes the Hilbert space of measurable, square-integrable 1-D function $f(x)$. For $f \in L^2(\mathbb{R})$, the wavelet transform of $f(x)$ at the scale a and position τ , is defined by

$$wf(a, t) = f * \psi_a(t) = \frac{1}{\sqrt{a}} \int f(x) \psi\left(\frac{x-t}{a}\right) dt \quad [1]$$

Where $*$ denotes the convolution operator, and

$$\psi_a(x) = \frac{1}{s} \psi\left(\frac{x}{s}\right) \quad [2]$$

The dynamic wavelet transform means the scale varies only along the dyadic sequence $(2^i)_{i \in \mathbb{Z}}$, and it has a fast algorithm. The first or second-order derivatives of a smoothing function $\theta(x)$ can be used to detect the sharp variation points. The extrema of the first derivative correspond to the edge points. $\theta(x, y)$ denotes a two dimension smoothing function. The first derivative respectively along the horizontal and vertical orientation is defined as basic wavelet function

$$\psi^1(x, y) = \frac{\partial \theta(x, y)}{\partial x} \quad \psi^2(x, y) = \frac{\partial \theta(x, y)}{\partial y} \quad [3]$$

The wavelet transform of $f(x, y)$ at scale a has two components defined by

$$W_a^1 f(x, y) = f * \psi_a^1(x, y) \quad \text{and}$$

$$W_a^2 f(x, y) = f * \psi_a^2(x, y)$$

Let a equal 2^j . At each scale, the modulus of the gradient vector is proportional to

$$M_{2^j} f(x, y) = \sqrt{\left| W_{2^j}^1 f(x, y) \right|^2 + \left| W_{2^j}^2 f(x, y) \right|^2}$$

The angle of the gradient vector with the horizontal orientation is given by

$$\text{Arg}[WTf(2^j, x, y)] = \text{tg}^{-1} \left[\frac{WT^2 f(2^j, x, y)}{WT^1 f(2^j, x, y)} \right]$$

So the edge points of $f * \theta_{2^j}(x, y)$ are the points (x, y) , where the modulus $M_{2^j} f(x, y)$ is local maxima in the direction of the gradient given by $A_{2^j} f(x, y)$. Many small local maxima are due to noise, so the maxima smaller than a given threshold value are eliminated. The basic wavelet in our algorithm is the quadratic spline wavelet, which is subject to the Battle-Lemarie spline wavelets family and is orthogonal with compact support and integral shift.

3.2.1 Fusion Algorithm of Multiscale Wavelet Transform

The multiscale edge images are obtained by first performing 2D multiscale Daubechies wavelet transform up to level two producing approximated, vertical, diagonal and horizontal details, which are important characteristics of an image. The 2D wavelet transform divides the image in to low frequency (L) and high frequency components (H) at level1. Further in level2 decomposition the low frequency information can be divided in to LL and high frequency information LH. The high frequency component in level1 is decomposed in to low frequency information HL and high frequency HH. The wavelet decomposition process is shown in the figure.5

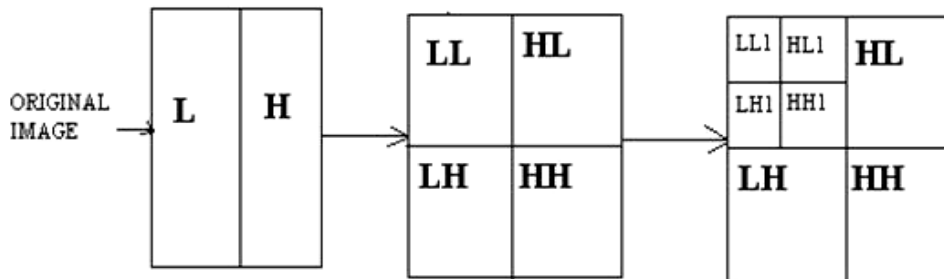


FIGURE 5: Wavelet Decomposition Based on Wavelet Transform

The results of wavelet decomposition using 2D Daubechies two wavelets for level2 on 'Cow' Video image frame is shown in figure6.

Though the wavelet transform applied on the image eliminates most of the noise in the image, the authentic edges in the image are also mixed with the noise, particularly the HH areas are greatly affected by noise. Thus many methods of edge detection discard these HH areas. This paper presents a new technique for edge detection by fusion algorithm based on wavelet transform and canny edge detector.

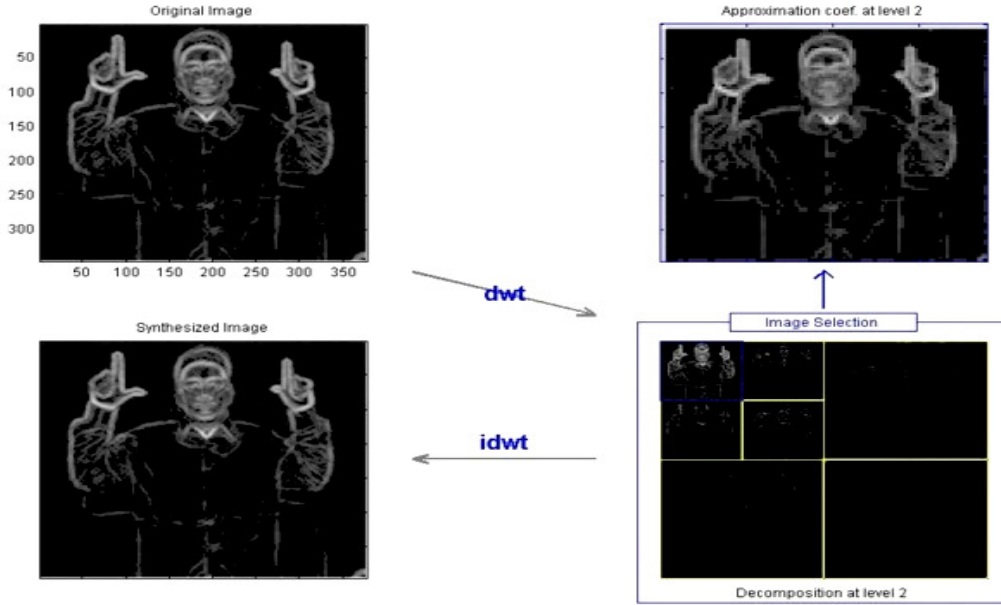


FIGURE 6: Daubechies2 wavelet transform of level2.

The notion $L^2(\mathbb{R}^2)$, where \mathbb{R} is a set of real numbers, denote the finite energy function $f(x,y)$ in \mathbb{R}^2 ; and x,y in \mathbb{R} . In two dimension wavelet transform, a 2D scaling function $\phi(x,y)$, and three two dimensional wavelets, $\phi^H(x,y)$, $\phi^V(x,y)$ and $\phi^D(x,y)$ are produced as shown in figure.5. The above functions represent gray level variations along different directions such as horizontal variations, vertical variations and diagonal variations. The DWT of $f(x,y)$ of size $M \times N$ is

$$W_\phi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f \phi_{j_0, m, n} \quad [4]$$

Where j, m, n, M, N are integers, $i = \{H, V, D\}$, j_0 is an arbitrary starting scale and the coefficients W_ϕ define an approximation of f at scale j_0 .

$$W_\phi^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f \psi_{j, m, n}^i \quad [5]$$

The coefficients in the above equation add horizontal, vertical and diagonal details as shown in figure.5 for scales $j \leq j_0$. The $\phi_{j, m, n}$ and $\psi_{j, m, n}^i$ denote scaled and translated basis functions as shown below:

$$\begin{aligned} \phi_{j, m, n}(x, y) &= 2^{j/2} \phi(2^j x - m, 2^j y - n) \\ \psi_{j, m, n}^i(x, y) &= 2^{j/2} \psi^i(2^j x - m, 2^j y - n) \end{aligned}$$

Given W_ϕ and W_ψ^i , f is obtained via inverse DWT as:

$$f = \frac{1}{\sqrt{MN}} \sum_m \sum_n \left(W_\phi^{j_0} \phi_{j_0} + \sum_i \sum_{j=j_0}^{\infty} W_\psi^j \psi_j^i \right) \quad [6]$$

Figure.7 shows the block diagram and figure 8 the application of image fusion process. Both the images undergo wavelet decomposition and reconstruction again in this stage using the above

equation. The fusion parameter selection rule is based on absolute maximum of horizontal coefficients and vertical coefficients:

$$W_{\varphi} = \begin{cases} W_{\psi}^H, & |W_{\psi}^H| \geq |W_{\psi}^V| \\ W_{\psi}^V, & \text{otherwise} \end{cases} \quad [7]$$

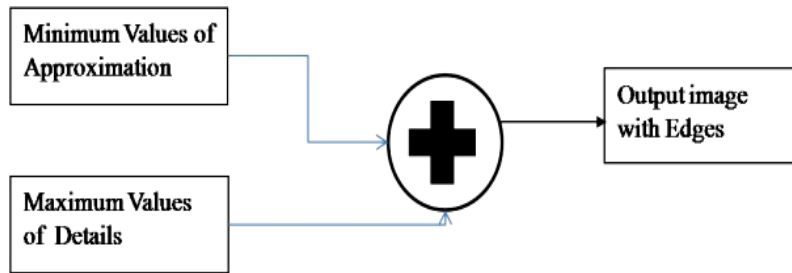


FIGURE.7. Image Fusion Process

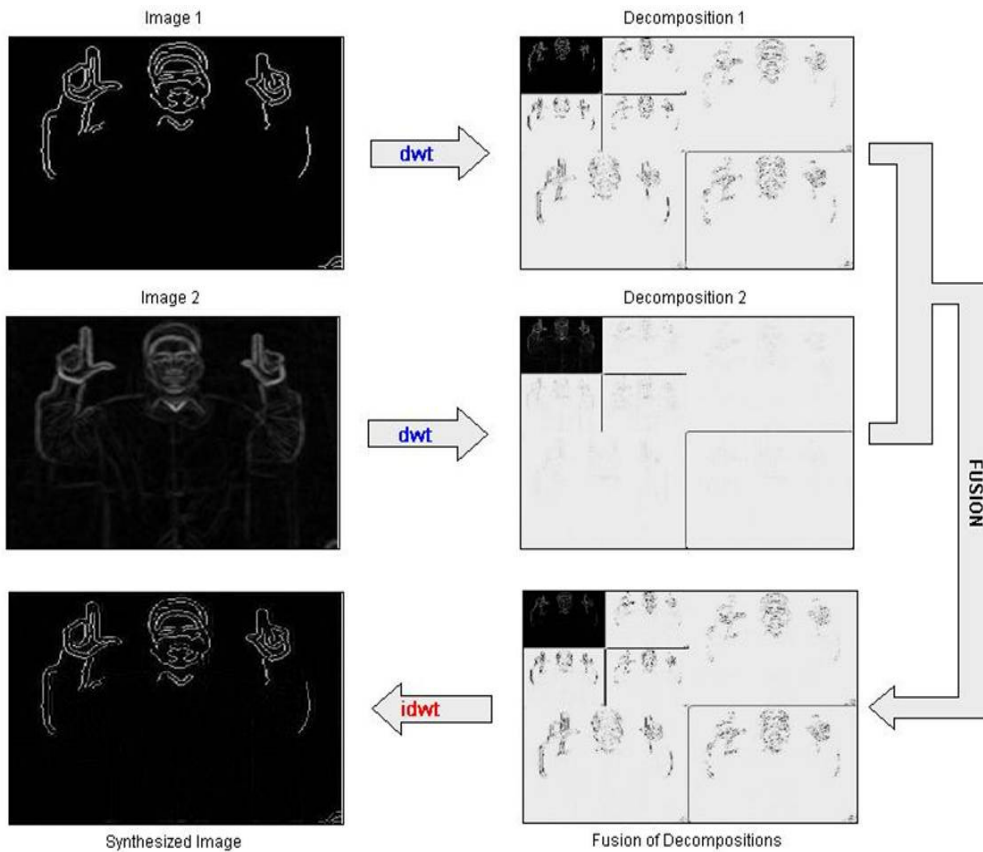


FIGURE 8: Image Fusion Applied to a video frame

To be acknowledged as a correct edge line, thresholding is desirable. Thresholding is performed on the wavelet reconstructed image by using a global thresholding algorithm known as Otsu's method. Converting a greyscale image to monochrome is a common image processing task. Otsu's method, named after its inventor Nobuyuki Otsu, is one of many binarization algorithms. Otsu's thresholding method involves iterating through all the possible threshold values

and calculating a measure of spread for the pixel levels each side of the threshold, i.e. the pixels that either falls in foreground or background. The aim is to find the threshold value where the sum of foreground and background spreads is at its minimum. In Otsu's method we exhaustively search for the threshold that minimizes the intra-class variance, defined as a weighted sum of variances of the two classes given by

$$\sigma_w^2(t) = W_1(t)\sigma_1^2(t) + W_2(t)\sigma_2^2(t) \quad [8]$$

Where W_i are probabilities of 'i' classes separated by a threshold 't' and σ^2 are variances of these classes. In this method image histogram is considered having a two Gaussian distribution representing the object and the background. This proves an effective algorithm for threshold calculation. After thresholding the video frames produced brighter edges compared to traditional algorithms for edge detection as shown in figure 9 for some frames of a sign.

3.3 Feature Extraction

High level feature extraction primary concern is finding shapes in computer images [32]. The feature extraction process is bound to shape our world view. During feature extraction process we seek invariance parameters so that the extraction process does not vary according to specified circumstances. That is, techniques used for feature extraction should find shapes reliably and robustly irrespective of changes in illumination levels, position, orientation and size of the object in a video. Objects in an image are represented as collection of pixels. For object recognition we need to describe the properties of these groups of pixels. The description of an object is a set of numbers called as object's descriptors.

Recognition is simply matching a set of shape descriptors from a set of known descriptors. A usable descriptor should possess four valuable properties. The descriptors should form a complete set, they should be congruent, rotation invariant and form a compact set. Objects in an image are characterized by two forms of descriptors region and shape descriptors. Region descriptors describe the arrangement of pixels within the object area whereas shape descriptors are the arrangement of pixels in the boundary of the object.

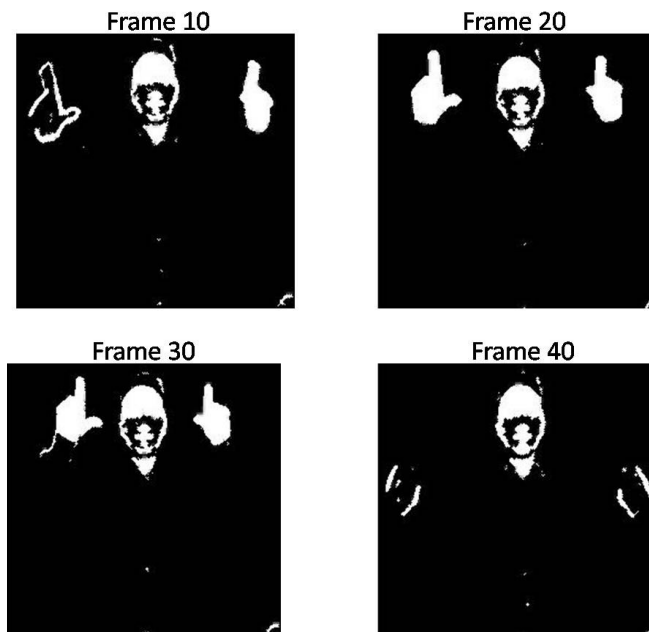


FIGURE 9: Result of Segmentation obtained by fusion of canny and DWT showing 4 frames in the video

3.3.1 Elliptical Fourier Descriptors

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 [32, 33, 34, 35]. 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 describe a curve in 2D space. Each pixel in the image plane is represented by a complex number. Thus any shape in an image is defined as

$$s(t) = x(t) + jy(t) \tag{9}$$

Where the parameter 't' is given by arc-length parameterization.

To obtain elliptical Fourier descriptors of a curve we need to obtain Fourier expansion of the shape in the above equation. The Trigonometric fourier expansion can be obtained using the expression

$$s(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(k\omega t) + b_k \sin(k\omega t)) \tag{10}$$

Accordingly the general expression for elliptical fourier coefficients for any closed contours can be defined as

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ c_0 \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \begin{bmatrix} \cos(kt) \\ \sin(kt) \end{bmatrix} \tag{11}$$

Where

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_0^{2\pi} x(t) dt & c_0 &= \frac{1}{2\pi} \int_0^{2\pi} y(t) dt \\ a_k &= \frac{1}{2\pi} \int_0^{2\pi} x(t) \cos(kt) dt & b_k &= \frac{1}{2\pi} \int_0^{2\pi} x(t) \sin(kt) dt \\ c_k &= \frac{1}{2\pi} \int_0^{2\pi} y(t) \cos(kt) dt & d_k &= \frac{1}{2\pi} \int_0^{2\pi} y(t) \sin(kt) dt \end{aligned}$$

Thus any closed contour can be represented by its fourier coefficients a_0, c_0, a_k, b_k, c_k and d_k . Where 'k' represents the rank of the ellipse and for $k=1$ corresponds to fundamental component of closed curve. For different values of k, the trigonometric summation defines the locus of an ellipse in a complex plane. As we change the parameter 't' the point traces ellipses moving at a speed proportional to the harmonic number 'k'.

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

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \times \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} \tag{12}$$

Where A is semi major axis oriented in horizontal axis direction and B is semi minor axis oriented in vertical axis direction. 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

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = E \times \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} \quad [13]$$

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 the defined parameters are as shown in the figure 10.

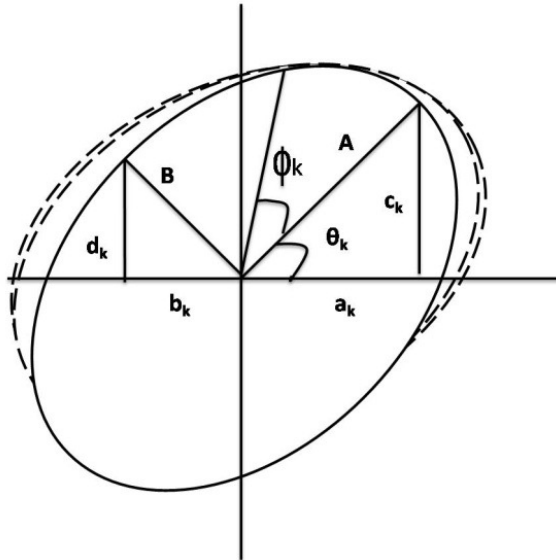


FIGURE 10: K^{th} Ellipse with Parameters.

Hence Eq. (13) becomes

$$\begin{aligned} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} &= \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \times \begin{bmatrix} \cos(kt) \\ \sin(kt) \end{bmatrix} \\ \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} &= \begin{bmatrix} A \cos(\theta) & -B \sin(\theta) \\ A \sin(\theta) & B \cos(\theta) \end{bmatrix} \times \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \\ &= E * \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} \end{aligned} \quad [14]$$

Where
$$E_k = \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix}$$

From the above equation we can write the fourier coefficients of k^{th} (a_k, b_k, c_k, d_k) as

$$\begin{aligned} a_k &= A_k \cos \theta_k \cos \phi_k - B_k \sin \theta_k \sin \phi_k \\ b_k &= -A_k \cos \theta_k \sin \phi_k - B_k \sin \theta_k \cos \phi_k \\ c_k &= A_k \sin \theta_k \cos \phi_k + B_k \cos \theta_k \sin \phi_k \\ d_k &= -A_k \sin \theta_k \sin \phi_k + B_k \cos \theta_k \cos \phi_k \end{aligned}$$

Where $A_k, B_k, \theta_k, \phi_k$ 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 figure 11.

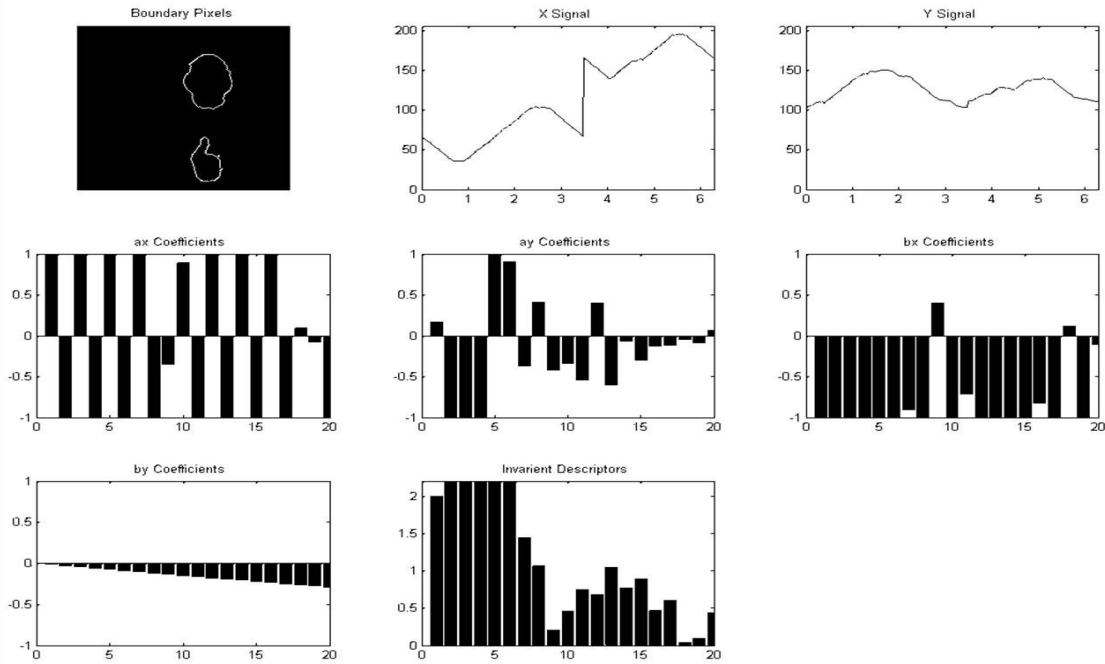


FIGURE 11: Plot gives the boundary pixels of a signers head and hand shapes represented using 20 coefficients and Fourier descriptors computed using direct implementation of Eq.11. The last figure shows the invariant descriptors.

The number of coefficients can be specified by the user by using the parameter 'k'. Details in the characterization are given by the number of coefficients used. For our application the maximum number of coefficients has been limited to 20. Figure 11 shows the original curve, $x(t)$ and $y(t)$ coordinate functions and the fourier descriptors described by invariant descriptors. The coefficients remain the same after changing the location, orientation and scale clearly indicating elliptical fourier descriptors truly characterize the shape of an human head and hands.

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.

3.3.2 Principal Component Analysis

Here we treat the current feature vector using principle component analysis (PCA) also known as karhunen Loeve(KL) transform 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. PCA reduces the dimensions of highly correlated input vectors. This technique has three effects: it orthogenesis the vector components obtained using fourier descriptors, secondly it orders the resulting orthogonal components in ascending order from largest variations to least variations and finally it discards the least contributing components of the data set.

The fourier descriptors matrix for a single video is 45x20 and we have 9 signers with around 80 signs each which makes our feature matrix huge. Now we reduce the dimension of our feature vector matrix by using PCA. Always start with normalizing the input vectors so they have zero mean and unity variance. The mathematics of PCA can be summarized in the following steps. Obtain the feature matrix \mathbf{c}_x from the data. Each column of the matrix defines a feature vector. Compute the covariance matrix Σ_x . This matrix gives information about the linear independence between the features. Obtain the Eigen values by solving the characteristic equation.

$$\det(\lambda_i I - \Sigma_x) = 0$$

These values form the diagonal covariance matrix Σ_y . Since the matrix is diagonal, each element is the variance of the transformed data.

Obtain the eigenvectors by solving for \mathbf{w}_i in

$$(\lambda_i I - \Sigma_x) \mathbf{w}_i = 0$$

for each Eigen value. Eigen- vectors should be normalized and linear independent. The transformation \mathbf{W} is obtained by considering the eigenvectors as their columns. Obtain the transform features by computing $\mathbf{c}_y = \mathbf{c}_x \mathbf{W}^T$. The new features are linearly independent. For classification applications, select the features with large values of λ_i . Remember that λ_i measures the variance and features that have large range of values will have large variance.

For example, two classification classes can be obtained by finding the mean value of the feature with largest λ_i .

For compression, reduce the dimensionality of the new feature vectors by setting to zero components with low λ_i values. Features in the original data space can be obtained by

$$\mathbf{C}_X^T = \mathbf{W}^T \mathbf{C}_Y^T$$

The result of elliptical Fourier descriptors and principle component analysis give a matrix consisting of feature vectors. Figure 12 shows feature vector matrix and figure 13 shows their uniqueness. Each row in the feature vector matrix corresponds to a video sign in the data base.

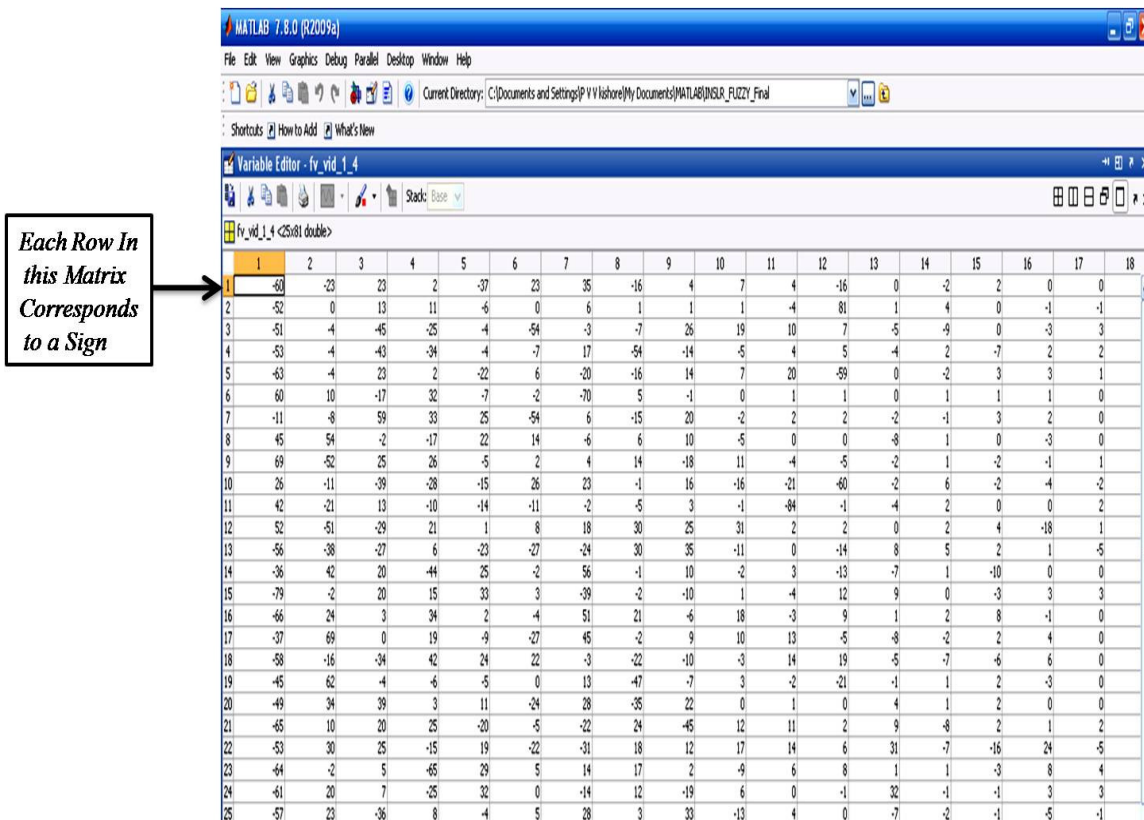


FIGURE 12: Feature Vector Matrix.

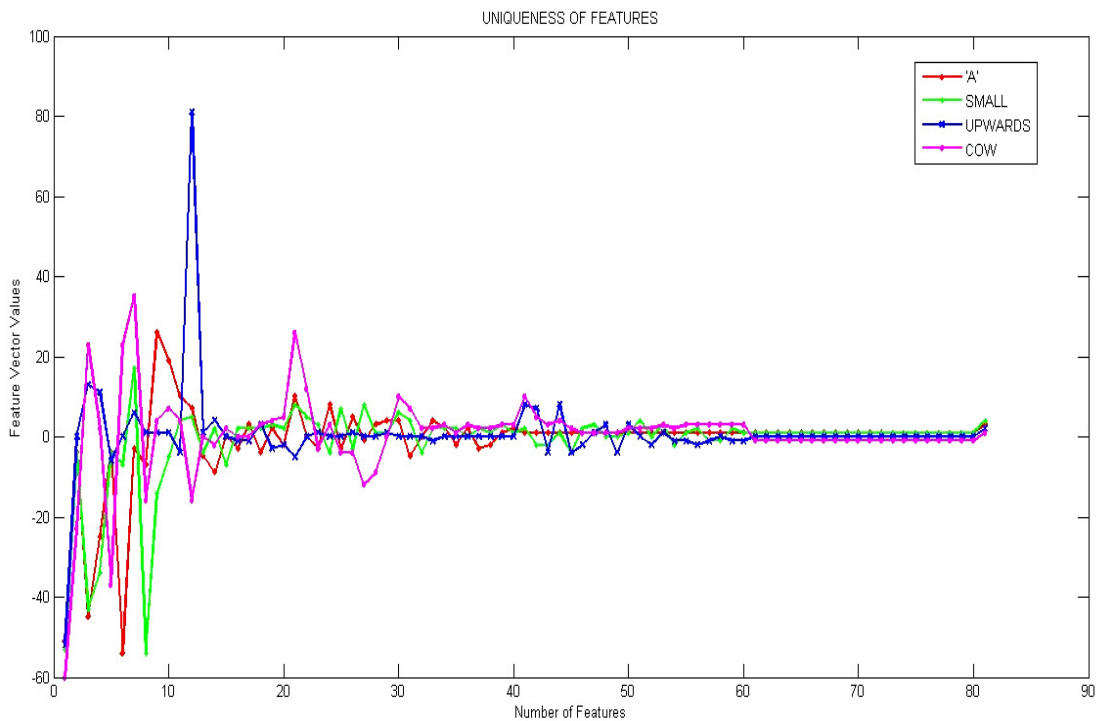


FIGURE 13: Plot showing uniqueness of feature vector for four different signs

4. PATTERN RECOGNITION

The final stage of the system is classification of different signs and generating voice messages corresponding to the correctly classified sign. In the past this part of the system was implemented using hidden markov models (HMM), Neural networks, Baye’s theorem and to some extent fuzzy logic. For our sign language recognition system we have deployed a fuzzy inference system.

4.1 Fuzzy Inference Systems (FIS)

Fuzzy logic is being applied to pattern recognition and pattern classification problems with if-then rules [36-37]. Numerous techniques have been proposed in the past to generate fuzzy if-then rules automatically [38-40]. There are various methods for generating fuzzy if-then rules for pattern classification problems. Generating fuzzy if-then rules from numerical data is considered for our system. The method was first proposed by Ishibuchi [41]. Generation of fuzzy if-then rules from numeric data is done in two sets of processes. One is mapping a pattern into fuzzy subspace using fuzzy partition and two is determination of fuzzy if-then rules each fuzzy subspace. An example of fuzzy partition system is given by a simple grid where a 2D pattern space is divided into fuzzy subspace.

The performance indicators in a if-then rule based fuzzy classification systems depends on the selection of fuzzy partition. If the fuzzy classification is too large, the performance may be low with many patterns being wrongly classified. If fuzzy partition is too small, many fuzzy if-then rules cannot be generated because lack of training patterns in the corresponding fuzzy subspaces. This problem was eliminated by the use of distributed fuzzy if-then rules [41] where all fuzzy if-then rules corresponding to several fuzzy subsets were simultaneously employed in fuzzy inference system. The main drawback of this system is that the number of fuzzy if-then rules becomes enormous for pattern classification problems.

4.2 Fuzzy Classification Method

For pattern classification we considered Takagi-sugeno-kang (TSK) or simply sugeno type Fuzzy inference system because the output membership functions are linear or constant. Sugeno fuzzy inference system consists of five 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 x_1 is Q_{1k} and x_2 is Q_{2k} and ... and x_n is Q_{nk} ,
THEN Z is

$$w = p_0 + p_1X_1 + p_2X_2 + \dots + p_nX_n \tag{15}$$

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

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

$$Q_{jk} = \exp \left[-\frac{1}{2} \left(\frac{X_j - \chi_{jk}}{\sigma} \right)^2 \right] \tag{16}$$

Where x_{jk} is the j^{th} input feature of x_k , the standard deviation σ of Gaussian MF given as

$$\sigma = \frac{1}{\sqrt{2a}} \tag{17}$$

5. RESULTS AND ANALYSIS

In this section we will analyze the performance of our system by its capability to recognize gestures from videos. We also discuss the difficulties faced while designing the system. Table 2 summarizes different gestures used for the analysis of our proposed system.

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 2: Summary of Sign language used

Classification of different gestures is done using Fuzzy inference system for 80 different gestures of Indian sign language by 9 different signers. Table 3 gives details of the fuzzy inference system used for gesture classification. Figure 14 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:

$$Recognition\ Rate\ (\%) = \frac{Number\ of\ Correctly\ Classified\ Signs}{Total\ Number\ of\ Signs} \times 100$$

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 3: Details of fuzzy inference system used for gesture classification.

A graphical User Interface (GUI) has been created to automatically train and recognize the gestures as shown in the figure 15. The GUI has two parts, a training part and a recognition part. Training is used to train new signs and store them in the data base of sign language system. Recognition tries to recognize the trained videos along with unknown videos of different signers. Training consists of four processes that can be activated by the use of buttons Load Video, Extract Features, Save Features along with the text in the text box and finally train button to finish the training. View FIS shows the user fuzzy inference system used. Once the load video button is pressed the video is played in the top portion of the training side which is interfaced to Microsoft windows media player.

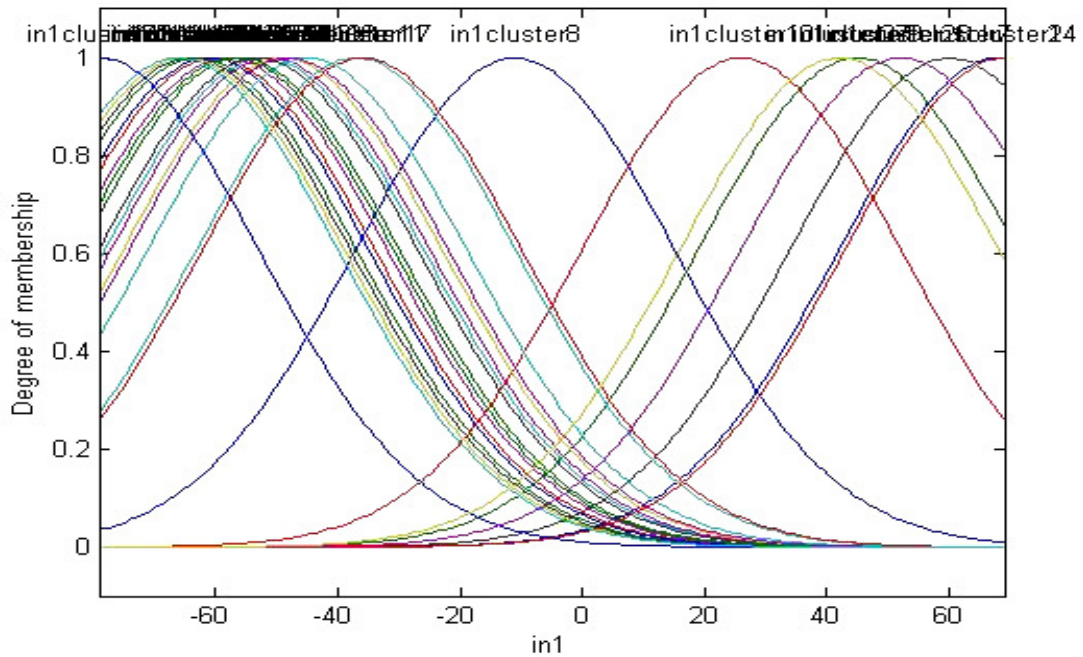


FIGURE 14: Input Membership Functions



FIGURE 15: GUI of Sign Language Recognition System

The recognition half consists of five buttons and a text box to display the word or sentence recognized by the system. The browse option is used to load video of the signer into the system. Once loaded and the recognize button is pressed, a voice message is heard along with the text message that is displayed in the text box provided on top of all buttons corresponding the gesture loaded into the system. The user can also see trained video list by using the trained videos list button. About ISLR button gives all the help information regarding our system.

Figure 16 gives a demo of training the fuzzy inference system using a video of signer representing a gesture of number 'six'. Figure 17 shows the recognition of sign 'six'. Table 4 shows the results obtained when training 10 samples for each gesture with different signers. Table shows recognition rates of some signs used for classification. 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.

6. CONCLUSION AND FUTURE WORK

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%. Most of the research in sign language is mainly the Extraction of Head and Hand Gesture Features for Recognition of Sign Language by the use of Sobel edge operator, DCT (Discrete Cosine Transform) for feature extraction and neural network for classification. In his framework he segmented head and hands separately and found the areas of the segmented head and hand portions. Then he applied DCT to extract features and finally a neural network is used to classify the gesturers.

When implemented the same process we observed that the sobel operator is unable to mark true edges. Hence we used canny operator with double thresholding to mark true edges on our videos. Though DCT based feature extraction gives minimum energy features but it fails to extract shapes of the segmented head and hand portions which was again done separately for all three areas increasing considerably the processing time of their algorithm. As we used Elliptical Fourier Descriptors which are proved shape extractors with minimum number of descriptors in our case it is 20 per frame. This greatly reduced processing time and extracted finite value feature vector.

Finally the video acquisition process was under controlled environment with good lighting and definite camera displacement. Our videos are shot under various lightning conditions with different signers as can be observed from Table 1.

Some researchers used hough transform to model hand shapes to extract features along with neural network to classify signs. Hough transform is good at identifying shapes but they should be almost straight lines. Hence the algorithm using hough transform was unable to give us good results for video of gesturers. It works well for images of gestures with only one hand at a time as we have identified. Hence we used Elliptical Fourier Descriptors which are proved shape extractors with minimum number of descriptors in our case it is 20 per frame. This greatly reduced processing time and extracted finite value feature vector.

Most authors used fuzzy logic for recognition of gestures as we did, but we observed that our algorithm gives better recognition rates for different gesturers under different conditions over a wide data base of signers. We also implemented the same process using error back propagation neural network giving fairly the same recognition rate.

In future we are looking at developing a system for Indian sign language that works in real-time. To accomplish this task we are building 3D models of hands and head of signer.

Sign	Correctly Recognized Signs	False Recogniti-on	Recognition Rate (%)
A	9	0	100
B	9	0	100
C	9	0	100
D	9	0	100
X	9	0	100
M	7	2	77.77
N	8	1	88.88
Y	6	3	66.67
Cow	8	1	88.89
Duck	9	0	100
Crow	5	4	55.56
Fat	8	1	88.89
Feather	7	2	77.77
Love	9	0	100
Together	9	0	100
Come here	6	3	66.67
Do your Home Work	5	4	55.56
Numbers 1-10	90	0	100
Upwards	9	0	100
Total	231	21	90.90%

TABLE 4: Results obtained when training signs with 10 different signers.



FIGURE 16: Training Demo of Gesture sign 'SIX'.



FIGURE 17: Recognition of sign 'SIX'

7. REFERENCES

- [1] W. Stokoe, D. Casterline, and C. Croneberg. "A Dictionary of American Sign Language on Linguistic Principles." Gallaudet College Press, Washington D.C., USA, 1965.
- [2] U. R. Wrobel. Referenz in eb"ardensprachen: Raum und Person. Forschungsberichte des Instituts f"ur Phonetik und Sprachliche Kommunikation der Universit"at M"unchen, 37:25–50, 2001.
- [3] C. Vogler and D. Metaxas. "A framework for recognizing the simultaneous aspects of american sign language". Computer Vision & Image Understanding, 81(3):358–384, March 2001.
- [4] Ruiduo Yang, Sudeep Sarkar, and Barbara Loeding. Enhanced level building algorithm to the movement epenthesis problem in sign language. In "CVPR", MN, USA, June 2007.
- [5] ASL corpus: <<http://www.bu.edu/asllrp/>>.
- [6] Christopoulos, C., Bonvillian, J., 1985. "Sign language". Journal of Communication Disorders 18, 1–20.
- [7] Atherton, M., 1999. Welsh today BSL tomorrow. In: Deaf Worlds 15 (1),pp. 11–15.
- [8] Engberg-Pedersen, E., 2003. From pointing to reference and predication: pointing signs, eyegaze, and head and body orientation in Danish Sign Language. In: Kita, Sotaro (Ed.),

- Pointing: Where Language, Culture, and Cognition Meet. Erlbaum, Mahwah, NJ, pp. 269–292.
- [9] Nyst, V., 2004. Verbs of motion in Adamorobe Sign Language. Poster. In: TISLR 8 Barcelona, September 30–October 2. Programme and Abstracts. (Internat. Conf. on Theoretical Issues in Sign Language Research; 8), pp. 127–129.
- [10] Abdel-Fattah, M.A., 2005. Arabic sign language: a perspective. *Journal of Deaf Studies and Deaf Education* 10 (2), 212–221.
- [11] Masataka, N. et al., 2006. Neural correlates for numerical processing in the manual mode. *Journal of Deaf Studies and Deaf Education* 11 (2), 144–152.
- [12] Indian Sign language, empowering the deaf, Indian Sign Language Database <<http://www.deafsigns.org>>.
- [13] Gaolin Fang and Wen Gao, “Large Vocabulary Continuous Sign language Recognition Based on Transition-Movement Models”, *IEEE Transaction on Systems, MAN, and Cybernetics-Vol.37, No.1, January 2007*, pp 1-9.
- [14] 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.375, 1995.
- [15] Ming-Hsuan Yang and Narendra Ahuja, “ Extraction of 2D Motion Trajectories and its Application to Hand Gesture Recognition”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.24, No.8, August 2002, pp1061-1074.
- [16] M.K.Bhuyan and P.K.Bora, “A Frame Work of Hand Gesture Recognition with Applications to Sigb Language”, *Annual India Conference, IEEE*, pp1-6.
- [17] Rini Akmeliawati, Melanie Po-Leen Ooi and Ye Chow Kuang , ‘Real-Time Malaysian Sign Language Translation Using Colour Segmentation and Neural Network’, *IEEE on Instrumentation and Measurement Technology Conference Proceeding, Warsaw, Poland 2006* ,pp. 1-6.
- [18] Nariman Habili, Cheng Chew Lim and Alireza Moini , ‘Segmentation Of The Face And Hands In Sign Language Video Sequences Using Color And Motion Cues’, *IEEE Transactions on Circuits and Systems For Video Technology 2004* , Vol. 14, No. 8, , pp.1086 – 1097
- [19] Yu Zhou and Xilin Chen, “Adaptive sign language recognition with Exemplar extraction and MAP/IVFS”, *IEEE signal processing letters*, Vol 17, No-3, March 2010, pp297-300.
- [20] Och J., Ney. H., 2002. Discriminative training and maximum entropy models for statistical machine translation. In: Annual Meeting of the Ass. For Computational Linguistics (ACL), Philadelphia, PA, pp. 295–302.
- [21] Sumita, E., Akiba, Y., Doi, T., et al., 2003. A Corpus-Centered Approach to Spoken Language Translation. Conf. of the Europ. Chapter of the Ass. For Computational Linguistics (EACL), Budapest, Hungary, pp. 171–174.
- [22] Casacuberta, F., Vidal, E., 2004. Machine translation with inferred stochastic finite-state transducers. *omputational Linguistics* 30 (2), 205–225.

- [23] Och, J., Ney, H., 2003. A systematic comparison of various alignment models. *Computational Linguistics* 29 (1), 19–51.
- [24] Koehn, P., 2004. Pharaoh: a beam search decoder for phrase-based statistical machine translation models. AMTA.
- [25] ESIGN project: <<http://www.sign-lang.uni-hamburg.de/eSIGN/>>.
- [26] ECHO corpus: <<http://www.let.kun.nl/sign-lang/echo/>>.
- [27] Liu Cai, “a kind of advanced Sobel image edge detection algorithm”, *Guizhou Industrial College Transaction (Natural Science Edition)*,2004, 33(5):77-79.
- [28] Baris Sumengen, B. S. Manjunath “ multi-scale edge detection and image segmentation”, ece dept. UC, santa Barbara, USA,(National Conference)2005.
- [29] Lily Rui, Liang, Carl G., Looney, “Competitive Fuzzy Edge Detection”, *Applied Soft Computing*, Vol.3, no.4, pp. 123-137, 2003.
- [30] G. Pajares, “A Wavelet-based Image Fusion Tutorial”, *Pattern Recognition*, Vol. 37, no.10, pp. 1855-1872, 2004.
- [31] J.N. Ellinas, M.S. Sangriotis, “Stereo Image Compression Using Wavelet Coefficients Morphology”, *Image and Vision Computing*, Vol.22, no.2, pp. 281-290, 2004.
- [32] Kuhl, F. P. and Giardina, C. R., *Elliptic Fourier Descriptors of a Closed Contour*, *CVGIP*, **18**,pp. 236–258, 1982
- [33] Lin C. C. and Chellappa, R., *Classification of Partial 2D Shapes using Fourier Descriptors*,*IEEE Trans. PAMI*, **9**(5), pp. 686–690, 1987.
- [34] Montiel, E., Aguado, A. S. and Zaluska, E., *Fourier Series Expansion of Irregular Curves*, *Fractals*, **5**(1), pp. 105–199, 1997.
- [35] Persoon, E. and Fu, K.-S., *Shape Description Using Fourier Descriptors*, *IEEE Trans. SMC*, **3**, pp. 170–179, 1977
- [36] [I] M. Sugeno, “An introductory survey of fuzzy control,” *Inform. Sci.*, vol. 36, pp. 59-83, 1985.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] H. Ishibuchi, K. Nozaki, and H. Tanaka, “Distributed representation of fuzzy rules and its application to pattern classification,” *Fuuy Sets and syst.*, vol.52, pp 21-32, 1992.