Indian Sign Language Recognition: An Approach Based on Fuzzy-Symbolic Data

Nagendraswamy H S[1] Chethana Kumara B M[2] and Lekha Chinmayi R[3]
Department of Studies in Computer Science
University of Mysore
Mysore, INDIA
[1]hsnswamy@compsci.uni-mysore.ac.in,[2]chethanbm.research@gmail.com,[3]lekha.2405@gmail.com

***Abstract*—In this paper, the task of recognizing signs made by hearing impaired people at sentence level has been addressed. A novel method of detecting sign boundaries in a video of continuous signs has been proposed and extraction of spatial features to capture hand movements of a signer through fuzzy membership functions has been proposed. Frames of a given video of a sign are preprocessed to extract face and hand components of a signer. The centroids of the extracted components are exploited to extract spatial features. The concept of interval valued type symbolic data has been explored to capture variations in the same sign made by the different signers at different instances of time. A suitable symbolic similarity measure is studied to establish matching between reference and test signs and a simple nearest neighbor classifier is used to recognize an unknown sign as one among the known signs by specifying a desired level of threshold. An extensive experimentation is conducted on a significantly large corpus of Indian regional signs created by us during the course of our research work in order to evaluate the performance of the proposed system.**

***Keywords*— Hearing impaired, Sign boundary, Key frame, Sign language, Fuzzy membership functions and Symbolic data.**

I. INTRODUCTION

Language is used a medium of communication to transmit information from one person to another. Sign language is used by hearing impaired people to express their thoughts and views with the society for their daily needs through hand gestures. Sign Language is a visual language, which consists of three major components such as finger-spelling- used to spell words letter by letter, word level sign vocabulary-used for the majority of communication and non-manual features-facial expressions and tongue, mouth and body position. Sign language is often misunderstood, as it is the signed version of the spoken languages such as English. Erroneously many people considered it as some coded form of spoken languages. However, it is a language by itself, with its own rule based grammar, symbol system and grammatical structure. Sign language differs from a spoken language in the sense that it is visual than auditory in nature and is composed of precise hand motions and shapes.

The communication between normal people and hearing impaired people might require an intermediate person, who can understand the sign language and interpret it in regular language and vice versa. However, this kind of communication not only makes the disabled person dependent on such professionals but also rather expensive in addition to not being practical for day-to-day communicative requirements of the people. A replacement of such professionals by a software tool for automatic recognition and translation of a sign language into a vocal language is needed in helping hearing impaired people to communicate with the society and also to help them to be independent up to certain level. Such an automated tool would acquire gestures, analyze, recognize and then produce equivalent sentences in a vocal language. Like a vocal language, sign languages are also evolved over time with several kinds of variations.

II. RELATED WORK

The sign language recognition research works have been addressed at finger spelling level in [2, 13, 14, 23, 26, 27], at word level [12, 17, 26] and at sentence level [4, 16, 17]. The techniques which gained importance due to their performance by research community are Ichetrichef moments [7], Gray level histogram [31], Sensor based glove technique [7,8,11,21], Hidden Morkov Models (HMM) [1], Hu moments and Electromyography (EMG) segmentation [1], Localized contour sequence [11], Size function [19], Transition-movement [6], Moment based size function [9], Convex chain coding and Basic chain code [21], Fourier descriptors [23], Grassman Covariance Matrix (GCM) [33], Fusion of appearance based and 5DT glove based features [21], Sparse Observation (SO) description [29].

From the literature survey, we have understood that the models proposed for sign language recognition address the problem either at finger spelling level [22, 26, 28, 29] or at word level [3, 10, 25, 32, 34]. Since signs used by hearing impaired people are very abstract, the sign language recognition based on fingerspelling or word seems to be cumbersome and not effective. With this observation, recently only two attempts were reported to address the problem at sentence level [4, 5,16, 17, 18]. Therefore, there is a scope for many more attempts in this direction.

In this research work, we have made an attempt to design a model to recognize signs of hearing impaired people at sentence level. We could explore the applicability of fuzzy-symbolic data for effective description and efficient representation of signs for their recognition.

III. PROPOSED METHODOLOGY

The proposed sign language recognition system involves five major phases namely, (A) Sign boundary detection (B) Segmentation of face and hands component from the frames of a given sign video (C) Extraction of relevant features through fuzzy theoretic approach (D) Compact representation of sign in the knowledgebase (E) Establishing matching between known and unknown signs for recognition.

A. Sign Boundary Detection (SBD) using Edge Orientation Histograms

A sign video normally consists of a signer performing many signs to communicate with other person(s). Effective analysis and interpretation of a given sign video essentially requires identification and segmentation of individual signs present in the video. This process is similar to the process of identification of shots in the field of video processing. In shot detection, a video is segmented into small meaningful chunks with each being a sequence of contiguous frames captured at a particular time interval without any breaks. Therefore, in general, shot boundary detection is the process of identifying the frames corresponding to boundary between two successive shots in the video.

A sign video with multiple signs is captured as a single shot by a camera continuously without any breaks. In human vision system, it is easy to identify the break points in a continuous sign. However, it is very difficult for a machine to learn and identify the break points. Hence, every signer was instructed to start and end every sign by a common hand gesture at a particular position, which can be treated as sign boundary. At the end of every sign, a boundary can be found similar to the shot boundary in normal videos, which helps us in separating multiple signs in a sign video. Hence, in this context, we call this process as sign boundary detection (SBD).

In this work, edge orientation histograms have been explored for detecting sign boundaries in a given sign video. Edges in a frame can characterize content of the frame very efficiently. In the image space an Edge orientation Histograms (EOH) represents the frequency and the directionality of the brightness variations. EOH is a unique feature, which cannot be duplicated by a color histogram or the homogeneous texture features. The edge orientation histograms act as convenient means of capturing and representing transitions in a continuous sign video as the movements of hands lead to a drastic change in the edge information of a sign. The edge orientations in a frame are represented in terms of 80 EOH features.

Initially, a sign video with multiple signs is converted into frames and transformed into grayscale for ease of processing. The frames in grayscale are then subjected to extract EOH features by applying edge detection using canny operator. Once the EOH features for all the frames are computed, we compare the first frame with all the other frames by computing Euclidean distances of EOHs corresponding to the frames. A histogram of the distances computed between first frame and all the other frames is created to visualize the boundaries of different signs in the given video. The Euclidean distance between the frames below a threshold are analyzed to arrive at the sign boundary locations in the histogram of distances. Fig.1. depicts the sequence of steps followed in sign boundary detection process.

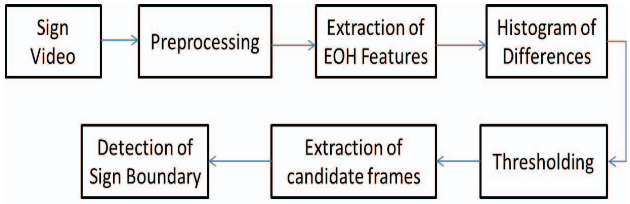


Fig. 1. Structural design of the proposed sign boundary detection method.

Fig.2. shows the illustration of EOH feature extraction process on two different frames. Table I presents an instance of EOH values computed for the frames shown in Fig.2.

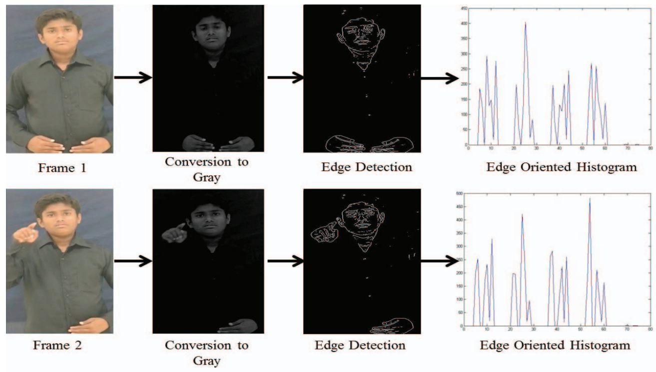


Fig. 2. Illustration of EOH feature extraction process on two different frames

TABLE I. EDGE ORIENTATION HISTOGRAM FEATURES FOR THE FRAMES SHOWN IN FIG.2.

Frame No.	F_1	F_2	F_3	F_4	…	F_{80}
1	0	56	50	66	…	0
2	0	197	253	0	…	0

As discussed earlier, every signer has followed same gesture both at the beginning and ending of every sign in a sign video. Hence, the frames close to beginning or close to ending of a sign video sequence with content similar to the first frame of a sign video will have smaller differences. This could be estimated using dissimilarity measure such as Euclidean distance and the computed distance values are used to plot a graph. The graph may show a series of crests and troughs, which depicts the hands orientation from initial to final frame. An empirically defined threshold was used to find the sign boundary. The distances, which are lesser than the threshold, are identified and the frames associated with those distances are considered as the candidate frames for sign boundary.

Let $F = \{ f_1, f_2, f_3, \ldots, f_n \}$ be the set of frames of a sign video and $F_t = \{ f_1, f_2, f_3, \ldots, f_k \} < t$ be the set of frames below the specified threshold. The frames having highest differences among the frames below the threshold will be considered and these frames represent the upwards and downwards hand movements up to certain level. Among the candidate frames, the frame having least distance will be called as sign boundary frame. If there is more than one frame having same least distance value then median frame will be considered as sign boundary. These frames categorize the different signs clubbed in a single sign video as individual signs. The distinguished individual signs are then exploited with various methodologies for sign language recognition system.

Fig.3. shows the histogram of differences between first frame and every other frame of a sign video. The red horizontal line represents the threshold for selecting frames to be analyzed. The red dots on the curve are the values of distances from first frame to all other frames below the threshold and the corresponding frames are treated as candidate frames. The green dots represent the candidate frames having highest distance below the threshold among all the candidate frames which we call as index frames for sign boundary identification.

Let S_i be the i^{th} sign in a given sign video, we can find two index frames corresponding to it say, IF_1^i and IF_2^i. Then, in order to detect the boundary between i^{th} and $(i+1)^{th}$ sign, we have to find a frame corresponding to the local minima among the candidate frames between the index frames IF_2^i and IF_1^{i+1}. For example, to compute the sign boundary between 1st and 2nd sign in a sign video, we have to identify a frame with least distance among all the candidate frames between the second index frame of the first sign IF_2^1 and first index frame of the second sign IF_1^2 which is represented by blue dots in the Fig.3.

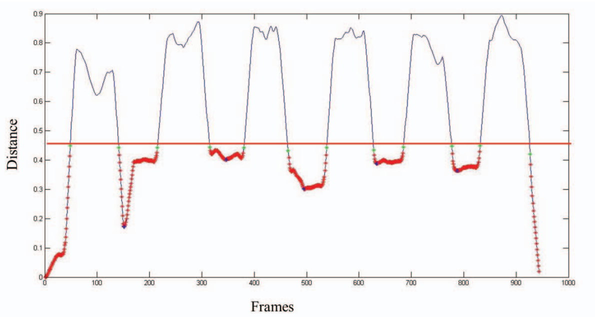


Fig. 3. Histogram of distances between first frame and remaining frames of a sign

B. Segmentation of Face and Hand Components

In this step, frames are extracted from the given sign video and are converted into YC_bC_r color space. Y is the luminance component and C_b and C_r are the blue-difference and red-difference chromium components. YC_bC_r color space provides more information for segmentation compared to RGB color space. Only luminance and blue-difference chromium values from YC_bC_r color space are considered to define a threshold value based on local information of a given frame in order to distinguish between skin and non-skin regions. Further, morphological operations such as opening and closing are used iteratively to get more accurately segmented components. Fig. 4. shows the process involved in segmenting face and hand components from the frames of a sign video. Fig. 5. shows the face and hand components before and after segmentation for two frames of two different signs as examples.

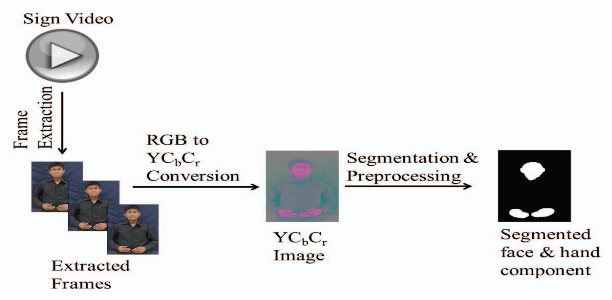
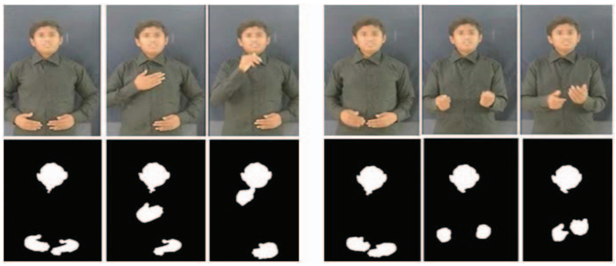


Fig. 4. Segmentation of face and hand components



a b

Fig. 5. Segmented face and hand components in various frames of the sentences a) *I WANT COFFEE* and b) *I WANT BUS TICKET*

C. Feature Extraction

Sign is comprised of three major components such as face, right hand and left hand. Face component is static in most of the signs and among two hands usually one hand is moved more compared to another hence it is called as manual hand whereas another hand, which is moved less is termed as non-manual hand.

We can observe that the position of signer's hands keep on changing across the frames while making a sign. Even though the face component is static most of the times, still we can observe some positional displacement. So it is possible to capture the position of these sign components across the frames of a sign video to define a signature and this signature can be used to characterize the sign. In view of this, we propose a method of capturing the positions of sign components in terms of fuzzy features by exploiting the concept of fuzzy membership function. Once the face and hand components are extracted from the frames, centroid of each component is computed. The computed centroid value is used to extract fuzzy features as follows.

The frames of sign video are uniformly divided into four components by drawing horizontal and vertical lines. Let a_x, b_x and c_x denote lower, middle and upper bound of the X-axis respectively. Similarly, let a_y, b_y and c_y represent lower, middle and upper bound of the Y-axis of the frame respectively. Then, the centroid $C(x_c,y_c)$ of a component in a given frame can be viewed in two directions, one from X-axis as reference and the other from Y-axis. Two spatial features μLx and μRx corresponds to X-axis and two spatial features

μUy and μLy corresponds to Y-axis are computed using fuzzy membership function as follows

$$(\mu Lx) = \begin{cases} 1 & if\,(a_x \le x_c \le b_x) \\ \dfrac{(x_c - a_x)}{(b_x - a_x)}, & if\,(c_x \ge x_c \ge b_x) \end{cases}$$

$$(\mu Rx) = \begin{cases} 1 & if\,(c_x \ge x_c \ge b_x) \\ \dfrac{(c_x - X_c)}{(c_x - b_x)}, & if\,(a_x \le x_c \le b_x) \end{cases}$$

$$(\mu Uy) = \begin{cases} 1 & if\,(a_y \le y_c \le b_y) \\ \dfrac{(y_c - a_y)}{(b_y - a_y)}, & if\,(c_y \ge Y_c \ge b_y) \end{cases}$$

$$(\mu Ly) = \begin{cases} 1 & if\,(c_y \ge y_c \ge b_y) \\ \dfrac{(c_y - y_c)}{(c_y - b_y)}, & if\,(a_y \le y_c \le b_y) \end{cases}$$

μLx denote the degree of belongingness of x coordinate of the centroid $C(x_c,y_c)$ to the first half of the frame along X-axis and μRx denote the degree of belongingness of x coordinate of the centroid $C(x_c,y_c)$ to the second half of the frame along X-axis.

Similarly, μUy denote the degree of belongingness of y coordinate of the centroid $C(x_c,y_c)$ to the first half of the frame along Y-axis and μLy denote the degree of belongingness of y coordinate of the centroid $C(x_c,y_c)$ to the second half of the frame along Y-axis.

Thus, every sign component is described with four spatial fuzzy features. Totally a frame having three components will be having 12 features. Fig. 6. shows an example instance of a frame horizontally and vertically along with the segmented sign components and their centroids. Table II presents the features extracted for frames with no overlapping of sign components.

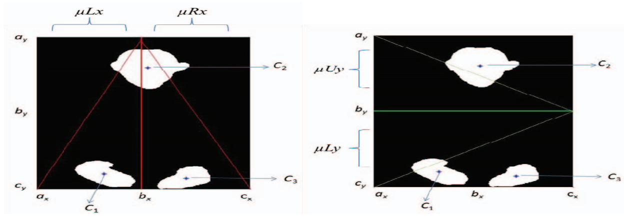


Fig. 6. Horizontal and Vertical split of a frame at midpoints b_x and b_y.

TABLE II. FEATURES EXTRACTED FOR GIVEN FRAMES

Frame	**Manual hand(C_1)**				**Face(C_2)**				**Non-manual Hand(C_3)**			
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1	1	0.637	0.190	1	0.963	1	1	0.400	0.599	1	0.133	1
2	1	0.637	0.191	1	0.961	1	1	0.400	0.598	1	0.134	1
3	1	0.637	0.192	1	0.959	1	1	0.399	0.598	1	0.134	1

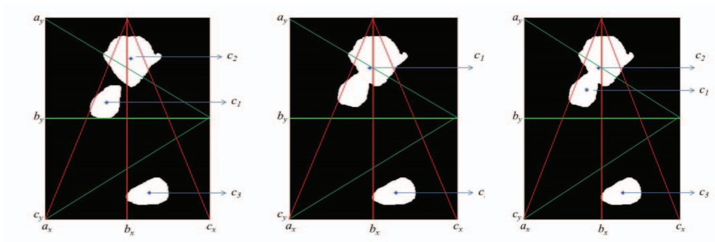
While making a sign, there are instances where the manual hand overlaps with the face component or with non-manual hand. In such cases, only two components are visible due to overlapping. In order to track the centroid of manual hand when there is a overlapping, we take the mean of the centroid of the manual hand in the previous frame (when there is no overlapping) and the centroid obtained from the overlapped manual hand and face component in the current frame.

Let (x_{ci}, y_{ci}) and (x_{ci+1}, y_{ci+1}) respectively denote the centroid of manual hand and the centroid of overlapped manual hand and face component in the ith and $(i+1)^{th}$ frame. The centroid for the manual hand in the $(i+1)^{th}$ frame (frame with overlapped manual hand and face component) is computed as follows

$$x c_{i+1}^{1} = \left(x c_i + x c_{i+1} \right) / 2$$

$$y c_{i+1}^{1} = \left(y c_i + y c_{i+1} \right) / 2$$

Since, the face component is almost static, in all the frames, the centroid computed for the face component in the first frame is considered for all the frames. Fig.7. shows centroids for frames with no overlapping, overlapped manual hand with face component and resolved frame.



(a) (b) (c)

Fig. 7. (a) The frame with no overlapping; (b) The frame with overlapping of manual hand with face component (c) The frame with computed centroid for the manual hand.

Similarly, when manual hand overlaps with non-manual hand in the $(i+1)^{th}$ frame, then the individual centroid for manual hand and non-manual hand is computed by taking the mean of their centroid computed for the overlapped manual and non-manual hand in the $(i+1)^{th}$ frame and their respective individual centroids computed in the i^{th} frame. Let (x_{ci1}, y_{ci1}) and (x_{ci2}, y_{ci2}) denote the centroid of the manual and non-manual hand in the ith frame. Let (x_{ci+1}, y_{ci+1}) denote the centroid of the overlapped manual and non-manual component in the $(i+1)^{th}$ frame. Then the centroid for manual and non-manual hand in the $(i+1)^{th}$ is computed as

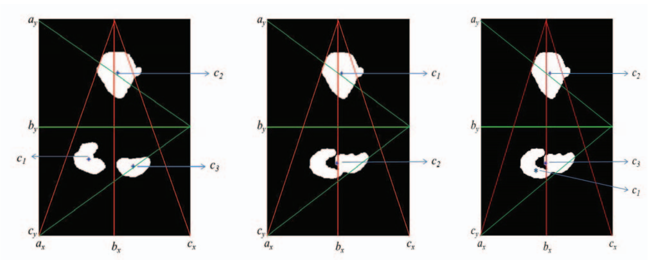
$$x_{c_{i+1,1}} = \left(x_{c_{i1}} + x_{c_{i+1}}\right)/2$$

$$y_{c_{i+1,1}} = \left(y_{c_{i1}} + y_{c_{i+1}}\right)/2$$

$$x_{c_{i+1,2}} = \left(x_{c_{i2}} + x_{c_{i+1}}\right)/2$$

$$y_{c_{i+1,2}} = \left(y_{c_{i2}} + y_{c_{i+1}}\right)/2$$

Once the centroids of all the three components are known, the features are extracted as discussed earlier. Fig.8 shows the illustration of the above process.



(a) (b) (c)

Fig. 8. (a) The frame with no overlapping; (b) The frame with overlapping of manual hand with non-manual hand (c) The frame with computed centroid for the manual hand.

D. Sign Representation

Major role in sign recognition system is to represent signs compactly in knowledgebase. In order to provide a robust representation for signs, issues such as unequal number of frames and intra-class variations must be addressed.

The variations in signs are due to different signers or same signer making same sign at different instances this may result in unequal number of frames due to their speed at which signs are made and captured. Perhaps, the adjacent frames in the sign video may not differ significantly in terms of their content. It is appropriate to remove such redundant frames, which do not differ significantly from previous frames. To obtain a fixed number of frames for each sign in knowledgebase, the idea of Key frame extraction is practiced.The concept of *K*-means clustering algorithm is explored in order to address the problem of selection of key frames from sign video. Conduction of sever experiments has led to choose a value for *K* empirically. After the *K* number of clusters is obtained for each sign, a representative frames for each cluster is chosen by looking into the frame having higher similarity with all the other frames in a cluster. Hence, *K* number of key frames characterizes every sign in knowledgebase.

Along with the problem of unequal number of frames, the instances of signs also contain some variations in terms of their content. Given sign instances are clusters to capture the intra-class variations. Once the clusters are obtained, multiple representatives are selected with the help of hierarchal cluster technique. In order to obtain the natural clusters while clustering the concept of Iinconsistency co-efficient (IC) is incorporated. IC values obtained depend on how the samples in a class are clustered at various levels. In order to find the adaptive threshold to cut the dendrogram to obtain natural clusters, we consider the maximum IC value where all samples are clustered into one cluster and the standard deviation (σ) of all the non-zero IC values obtained. The threshold (TH) is computed by subtracting the standard deviation (σ) multiplied by a small constant value (β) from the maximum IC values as follows:

$$\text{TH} = \max(\text{IC}) - (\beta * \sigma) \quad (1)$$

Through numerous experiments, β values are empirically chosen. For the purpose of training and testing we consider different percentages of samples (60:40, 50:50, 40:60) within the obtained cluster of various instances of a particular sign. The interval valued feature vectors which represent entire cluster is formed by aggregating the feature vectors of training and testing samples which represents each instance in a cluster. Minimum and maximum feature values are found using min( ) and max( ) operations to construct interval valued type of feature vector. The process of deriving interval valued type symbolic feature vector for a cluster is described as follows:

Let $S_1, S_2, S_3, ..., S_n$ be the n number of signs considered by the system for study and let $S_i = \{s_1, s_2, s_3, ..., s_m\}$ be the m number of instances of a sign S_i made by the signers at different instances of time.

Let $\{KF_{i1}, KF_{i2}, KF_{i3}, ..., KF_{it}\}$ be the *'t'* number of key frames chosen for the video of one of the instances of a sign S_i, where $KF_{ij} = \{f_{ij}^1, f_{ij}^2, f_{ij}^3, ..., f_{ij}^l\}$ be the feature vector representing jth key frame of one of the instances of a sign S_i, and l is the number of features.

Let d be the number of clusters obtained from m instances of a sign S_i. If a particular cluster say, p among c number of clusters containing q number of instances, then the features describing the j^{th} key frame of all the q number of instances are aggregated to form an interval type symbolic data as described below

Let

$$KF_{ij}^{(1)} = \{KF_{ij}^{(1)1}, KF_{ij}^{(1)2}, KF_{ij}^{(1)3}, ..., KF_{ij}^{(1)l}\}$$
$$KF_{ij}^{(2)} = \{KF_{ij}^{(2)1}, KF_{ij}^{(2)2}, KF_{ij}^{(2)3}, ..., KF_{ij}^{(2)l}\}$$
$$KF_{ij}^{(3)} = \{KF_{ij}^{(3)1}, KF_{ij}^{(3)2}, KF_{ij}^{(3)3}, ..., KF_{ij}^{(3)l}\}$$
$$\vdots$$
$$KF_{ij}^{(q)} = \{KF_{ij}^{(q)1}, KF_{ij}^{(q)2}, KF_{ij}^{(q)3}, ..., KF_{ij}^{(q)l}\}$$

be the feature vectors representing the j^{th} key frame of the 1^{st}, 2^{nd}, 3^{rd},…, q^{th} instances of a cluster p, respectively. Then,

$$f_{ij}^{1-} = Min\left\{f_{ij}^{(1)1}, f_{ij}^{(2)1}, f_{ij}^{(3)1}, \ldots, f_{ij}^{(q)l}\right\}$$

$$f_{ij}^{1+} = Max\left\{f_{ij}^{(1)1}, f_{ij}^{(2)1}, f_{ij}^{(3)1}, \ldots, f_{ij}^{(q)l}\right\}$$

Similarly, we compute

$$f_{ij}^{2-},\ f_{ij}^{2+},\ f_{ij}^{3-},\ f_{ij}^{3+},\ \ldots.,\ f_{ij}^{l-},\ f_{ij}^{l+}$$

Thus, the aggregated j^{th} key frame of reference feature vector representing the p^{th} cluster of a sign S_i is given by,

$$RF_{ij}^{p} = \left\{\left[f_{ij}^{(p)1-}, f_{ij}^{(p)1+}\right], \left[f_{ij}^{(p)2-}, f_{ij}^{(p)2+}\right], \left[f_{ij}^{(p)3-}, f_{ij}^{(p)3+}\right], \ldots, \left[f_{ij}^{(p)l-}, f_{ij}^{(p)l+}\right]\right\}$$

E. Matching and Recognition

In the previous section we discussed how features are extracted and represented. The recognition task is accomplished by the comparison between test feature vector and the entire reference sign feature vector stored in knowledgebase.

In this phenomenon a similarity value is computed and the reference sign having maximum similarity value with the test sign is considered. The comparison between reference sign feature vector and the test sign feature vector is carried out by using the similarity measure proposed in [15] which is as follows:

Let $TF_j = \left\{f_j^1, f_j^2, f_j^3, \ldots, f_j^l\right\}$ be the crisp feature vector and $RF_j = \left\{[f_j^{1-}, f_j^{1+}], [f_j^{2-}, f_j^{2+}], [f_j^{3-}, f_j^{3+}], \ldots, [f_j^{l-}, f_j^{l+}]\right\}$ be the interval valued type symbolic feature vector as representing the j^{th} key frame of a test and a reference sign respectively. Similarity between the test and reference sign with respect to j^{th} key frame is computed as

$$SIM(RF_j, TF_j) \frac{1}{l}\sum_{d=1}^{l}\left\{\begin{array}{ll} 1 & if\,(f_j^{d-} \le f_j^{d} \le f_j^{d+}) \\ \max\left[\dfrac{1}{1+abs(f_j^{d-}-f_j^{d})}, \dfrac{1}{1+abs(f_j^{d+}-f_j^{d})}\right] & otherwise \end{array}\right\} \quad (2)$$

The total similarity between the test and reference sign due to all the frames is computed as

$$SIM(RF, TF) = \sum_{j=1}^{L} SIM(RF_j, TF_j) \quad (3)$$

The number of frames in test and reference sign is denoted by L which is 40. The given test sign is recognized as one among the known sign in the sign knowledgebase by incorporating the perception of nearest neighbor classification methodology.

IV. EXPERIMENTATION

In order to demonstrate the efficacy of the proposed method, we have conducted experimentations on UoM-ISL sign language dataset. The sentences used in day to day life by hearing impaired people are described in our dataset. The dataset is consisting of 1040 (17.3 hours) sign videos of 26 diverse signs articulated by four signers with ten repetitions. The sign videos performed by hearing impaired students of various schools of Mysuru region has been considered in our experiments.

The values for β in Eqn. (1) can be in the range 0.1 to 1.0. However, the number of clusters is significantly changed only for few values of β. In our experiments, we institute that the number of clusters obtained for β = (1, 0.5, 0.1) are significantly diverse and hence, we chose these three values for three different experiments on the dataset of signs considered.

Numerous experiments are carried out for different percentages of training and testing samples (60:40, 50:50 and 40:60). Random sets of training and testing samples are chosen from each cluster for repeated experiments. The performance of the system is measured in terms of F-measure in each experiment.

Table III gives the recognition rate of all 50 random runs for each class for various percentages of training and testing samples for one of the three different numbers of clusters (382 representatives).

TABLE III. Average Recognition rate of the proposed method for 382 representatives

Class Index	Training : Testing			Class Index	Training : Testing		
	40:60	50:50	60:40		40:60	50:50	60:40
1	0.90	0.93	0.97	14	0.86	0.93	0.97
2	1.00	0.97	0.94	15	0.86	0.89	0.77
3	0.67	0.73	0.85	16	0.71	0.52	0.49
4	0.93	0.93	0.97	17	0.80	0.71	0.62
5	0.80	0.96	0.92	18	0.75	0.88	0.73
6	0.78	0.67	0.79	19	0.90	0.81	0.78
7	0.56	0.74	0.80	20	0.73	0.79	0.81
8	0.88	0.91	0.94	21	0.81	0.89	0.92
9	0.64	0.77	0.69	22	0.87	0.77	0.90
10	0.73	0.88	0.88	23	0.76	0.69	0.84
11	0.79	1.00	0.97	24	0.75	0.81	0.81
12	1.00	1.00	1.00	25	0.89	0.52	0.64
13	0.83	0.69	0.74	26	0.93	0.93	0.93

Table IV gives the recognition performance of the proposed method in terms of F-measure for various percentages of training and testing samples for the complete database and for three different numbers of clusters.

TABLE IV. RECOGNITION PERFORMANCE OF THE PROPOSED METHOD IN TERMS OF F-MEASURE FOR DIFFERENT PERCENTAGES OF TRAINING AND TESTING SAMPLES FOR FUZZY LOGIC APPROACH

Sign Language Data Set	Ratio of Training and Testing	Number of representatives	Overall Average Recognition Rate
UoM-ISL Data	60:40	273	76.02±0.99
		308	79.72±0.71
		382	**84.82±1.05**
	50:50	273	75.15±0.27
		308	80.83±0.98
		382	**82.05±0.70**
	40:60	273	74.81±0.86
		308	77.61±0.71
		382	**79.81±0.69**

In the proposed system with symbolic representation technique, out of 1040 signs, 658 signs selected through clustering process are used as reference signs in the knowledgebase and the remaining 382 samples are used for testing the system (for 60:40 training and testing).

When compared to other validation techniques, the proposed system uses less number of reference signs in the knowledgebase. Thus, we claim that the proposed method is efficient in terms of storage and robust in terms of representation as it effectively captures the intra-class variations in the form of interval type data. Fig.9 to Fig.11. delivers the confusion matrices for (50:50), (60:40) and (40:60) training and testing samples with 382 sign representatives respectively.

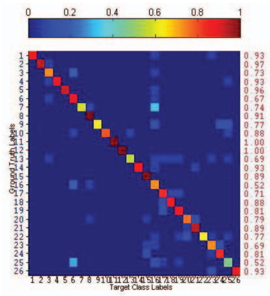


Fig. 9. The confusion matrix for (50:50) training and testing samples with 382 sign representatives for fuzzy logic based approach

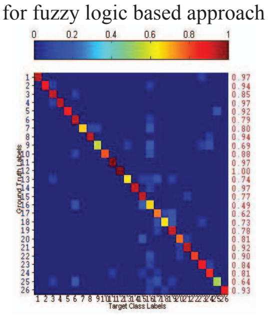


Fig. 10. The confusion matrix for (60:40) training and testing samples with 382 sign representatives for fuzzy logic based approach

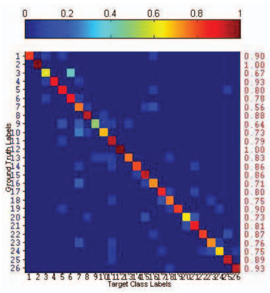


Fig. 11. The confusion matrices for (50:50), (60:40) and (40:60) training and testing samples with 382 sign representatives respectively.

Experiments were also conducted to study the performance of the proposed system for signers' independent sign recognition. Out of four signers, signs made by three signers are used to train the system and the signs made by the other signer are used for testing and the results are shown in Table V. The same experiment is repeated four times, where at each time signs made by one of the signers are used for testing while the signs made by the other three signers are used to train the system.

TABLE V. PERFORMANCE OF THE PROPOSED METHOD FOR DIFFERENT PERCENTAGES OF TRAINING AND RESTING SAMPLES FOR SIGNER INDEPENDENT SIGN RECOGNITION

Sign Language Data Set	**Average recognition rate**		
UoM-ISL Data	Training	Testing	Average F-measure
	2,3,4	1	58.30±1.50
	1,3,4	2	59.49±0.77
	1,2,4	3	58.90±1.00
	1,2,3	4	59.07±0.59
Overall average recognition rate			58.94

Table VI shows the performance of the proposed method using Leave-one-out and 10 fold Cross validation. Thus, a leave-one-out classification technique has been followed in this experiment. Thus, out of 1040 signs, 780 signs are used for training and 260 signs for testing. Performance of the system has been measured in terms of F-measure for the entire database and the results are presented in Table VI.

TABLE VI. RESULTS OBTAINED BY THE PROPOSED METHODOLOGY FOR DIFFERENT VALIDATION TECHNIQUES

Different Validation techniques	**Number of reference signs stored in knowledgebase**	**Recognition rate (Average F-measure)**
Leave-one-out	1014	73.04 ± 0.81
10 fold Cross validation	780	73.96 ± 0.79

V. CONCLUSION

In this work, the concept of fuzzy trapezoidal membership function has been explored to characterize the sign effectively. Symbolic data analysis particularly the interval valued type representation is explored to effectively capture the intra-class variations of signs. Applicability of symbolic similarity measure is studied for establishing effective matching of signs for recognition. Performance of this approach has been evaluated in terms of average F-measure through the conduction of extensive experimentation on newly created UoM-ISL dataset. The proposed system yields the recognition rate of 84.82% for 382 representatives. We can infer that as the number of representatives is directly proportional to the recognition rate of the system. However, the scalability of the proposed method needs to be studied on a large dataset. The proposed method has been compared with state-of-the-art validation techniques and the results found to be superior in terms of recognition accuracy and storage requirements

REFERENCES

[1] M. Al-Ahdal, and N. Tahir, "Review in sign language recognition systems." *IEEE Symposium on Computers and Informatics (ISCI)*, 2012, pp.52-57.

[2] Archana S. Ghotkar and G K. Kharate, "Study of Vision Based Hand Gesture Recognition using Indian sign language." *IJSS and Intelligent Systems*, March 2014, vol.7, No. 1.

[3] Brashear, Helene, et al., "Using multiple sensors for mobile sign language recognition." *Georgia Institute of Technology*, 2003.

[4] B.M. Chethana Kumara and H. S. Nagendraswamy. "Sign Energy Images for Recognition of Sign Language at Sentence Level." 2016 Apr; 139(2).

[5] B. M. Chethana Kumara, H. S. Nagendraswamy and R. Lekha Chinmayi, "Spatial Relationship Based Features for Indian Sign Language Recognition." *Int'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE)*, 2016, Vol. 3, Issue 2, ISSN 2349-1469 EISSN 2349-1477.

[6] Fang, Gaolin, Wen Gao, and Debin Zhao, "Large-vocabulary continuous sign language recognition based on transition-movement models." *Systems, Man and Cybernetics, Part A: Systems and Humans*, IEEE Transactions, 2007, 37.1, 1-9.

[7] C. Gourley, "Neural Network Utilizing Posture Input for Sign Language Recognition." *Technical report computer vision and robotics research laboratory*, University of Tenessee Knoxville, November 1994.

[8] M. Handouyahia, D. Zion, and S. Wang, "Sign Language Recognition using moment based size functions." *Vision interface'99, Trois-Rivieres, Canada*, May, 1999, pp19-21.

[9] M. Handouyahia, "Sign Language Recognition using moment based size functions." *MSc en Informatique demathematique et d informations*, universite de sherbrooke, sherbrooke, 1998.

[10] Hanjie Wang., Xinjnan Chai., Xilin Chan. Sparse Observation Alignment for Sign Language Recognition." *Neurocomputing*, 175, 2016, pp 674-685.

[11] W.W. Kong and Surendra Ranganath, "Towards subject independent continuous sign language recognition: A segment and merge approach. " *Pattern Recognition*, 47, 2014, 1294–1308.

[12] Liwicki, Stephan, and Mark Everingham, "Automatic recognition of finger spelled words in British sign language." *Computer Vision and Pattern Recognition Workshops, CVPR Workshops, IEEE Computer Society Conference. IEEE*, 2009.

[13] D. McNeill, "Hand and mind: What gesture revel about thought." Chicago.IL: University of Chicago Press, 1992.

[14] K. Murakami and H. Taguchi, "Gesture Recognition using Recurrent Neural Network." *Actes de CHI'91 Workshop on user interfaces by Hand Gesture, ACM*, 1991, pp 237-242.

[15] H. S. Nagendraswamy and D. S. Guru, "A New Method of Representing and Matching Two Dimensional Shapes." *Int. J. Image Graphics* 7(2): 2007, pp377-405.

[16] Nagendraswamy, H. S., BM Chethana Kumara, and R. Lekha Chinmayi, "GIST Descriptors for Sign Language Recognition: An Approach Based on Symbolic Representation." *MIKE, Springer International Publishing*, 2015, pp 103-114.

[17] H. S. Nagendraswamy, B. M. Chethana Kumara, D. S. Guru, and Y. G. Naresh, "Symbolic Representation of Sign Language at Sentence Level." *Int . Jl. of Image and Graphics Signal Processing* 7.9 2015: 49.

[18] H. S. Nagendraswamy, and B. M. Chethana Kumara, "LBPV for Recognition of Sign Language at Sentence Level: An Approach Based on Symbolic Representation." *Journal of Intelligent Systems. Journal of Intelligent Systems*. April 2016, ISSN (Print) 0334-1860, DOI: 10.1515/jisys-2015-0003.

[19] Pentland Holden, Eun-Jung, Gareth Lee, and Robyn Owens., "Australian sign language recognition." *Machine Vision and Application,* 16.5 2005, 312-320, 2005.

[20] M. Punchimudiyanse, and R.G.N. Meegama, "3D animation frame work for Sign Language." *In the proceedings of ICET*, Colombo 2015.

[21] Ramakant, N.-E.-K. Shaik and L. Veerapalli," Sign language recognition through fusion of 5DT data glove and camera based information." *Advance Computing Conference (IACC)*, 2015, pp.639-643, 12-13 June 2015.

[22] J. Rekha, J. Bhattacharya, and S. Majumder. "Shape, texture and local movement hand gesture features for Indian sign language recognition." *Trendz in Information Sciences and Computing (TISC), 3rd International Conference,* IEEE, 2011.

[23] A. M. Riad, H. K. Elmonier, S. Shohieb, & A. S. Asem, "SIGNWORLD; Deeping into the silence world and hearing its signs (State of the Art)." 2012, *ArXiv preprint arXiv*:1203.4176.

[24] Robert M Haralick and Linda G Shaipro, "Local invariant feature detectors: A Survey." *Computer and Robot vision* Vol.2, Addison –Wesley publishing company, 1993.

[25] T. Starner, "Visual Recognition of American Sign Language Using Hidden Markov Models." *Master Thesis, MIT, Media laboratory*. Feb, 1995.

[26] Starner, Thad, Joshua Weaver, and Alex Pentland, Real-time American sign language recognition using desk and wearable computer based video. *Pattern Analysis and Machine Intelligence, IEEE Transactions* 20.12, 1998, pp 1371-1375.

[27] M. G. Suraj and D. S. Guru, Appearance Based Recognition Methodology for Recognizing Fingerspelling Alphabets. *IJCAI* 2007, pp 605-610.

[28] M.G. Suraj and D.S. Guru, "Secondary Diagonal FLD for Fingerspelling Recognition." *ICCTA,International Conference on Computing: Theory and Applications, International Conference on Computing: Theory and Applications*, 2007, pp. 693-697,

[29] Wang Chunli, Wen Gao, and Shiguang Shan, "An approach based on phonemes to large vocabulary Chinese sign language recognition." *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on. IEEE*.

[30] Wen Gao, Enhanced user by hand gesture recognition." *CHI'95 workshop on user interface by hand gesture*, Denver, 1995, pp 45-53.

[31] T. William, Freeman and Michal Roth, "Orientation Histograms for Hand Gesture Recognition." *Int. Workshop on Automatic Face and Gesture Recognition*, Switzerland, 1995.

[32] Wu Jiangqin, Gao wen, Song yibo, Liu wei and Pang bo, " A simple sign language recognition system based on data glove." *Fourth International Conference on Signal Processing Proceedings*, 1998.

[33] Xiujunan Chai, Hanjie Wang, Fang Yin and Xilin Chen., "Communication Tool for the Hard of Hearings. A Large Vocabulary Sign Language Recognition System." *International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015.

[34] Youssif, A. A. Aliaa, Amal Elsayed Aboutabl, and Heba Hamdy Ali, "Arabic sign language recognition system using HMM." *IJACSA*, 2011, 2.11.