VIDEO BASED INDIAN SIGN LANGUAGE RECOGNITION USING BLOCK ZIG-ZAG DCT FEATURES AND MAHALANOBIS DISTANCE CLASSIFIER

Sunita Ravi[1,2], M. Suman[3] and P. V. V. Kishore[1]
[1]Department of Electronics and Communication Engineering, K. L. University, Green Fields, Vaddeswaram, Guntur DT, India
[2]Department of Electronics and Communication Engineering, NRI Institute of Technology, India
[3]Department of Electronics and Computer Engineering, K. L. University, Green Fields, Vaddeswaram, Guntur DT, India
E-Mail: Sunita71ravi@gmail.com

ABSTRACT

The objective of this paper is to recognize discrete words from videos of Indian sign language. Sign language recognition is still not popular research area in India till recently. This paper introduces a fast model to extract hands form the video sequences and to generate features. We introduce a local cosine feature (LCF) to describe the hand shapes with minimum number of features. It is based on 2D discrete cosine transform (DCT) applied through total variational model. The features for each sign video are classified with Mahalanobis distance classifier. Mahalanobis distance based pattern analysis is becoming popular due to their compactness and faster time to execution. A total of 20 isolated words are from Indian Sign Language (ISL) are trained and tested. Experimental results using the proposed model produced a recognition rate of 90.44%, which when compared to system with DCT features which is 81%.

Keywords: Indian sign language recognition, hand signs, discrete cosine transform, support vector machines, recognition rate.

1. INTRODUCTION

In this paper, we address the problem of sign language recognition for Indian sign language. Sign language is a visual language of the hearing impaired. The language is defined by finger shapes, movements, hand movements and head movements. Most of the signs focus on hand shapes.

The main challenge in any sign language recognition system is to find a computer model that can fully capture the large-scale vocabularies of the language. Hence the necessity to develop a Graphical User Interface called Indian Sign Language Recognition (In SLR-GUI) that can be used to train and test a large database of sign language vocabulary with less memory usage.

We started by extracting signer hands and head shapes on video frames captured under simple backgrounds. Simple backgrounds in the sense that the signer should wear a dark colored full sleeves jacket under dark background as shown in the Table-1.

The sign videos are pre-processed to reduce the size [25], remove noise during capturing [26] and convert to a gray scale image [25] to facilitate the ease of further processing. Sign video segmentation [27-30] is performed to extract hand shapes and head portion. Sobel edge operator [31] is applied to extract edges and thus segment the required hand and head portions. Hands and head segments are separated into individual segments with repeated dilation and erosion operations providing hand regions [32, 33]. Feature extraction is through Local Cosine Features (LCF) on hand regions. LCF are 2D DCT on an 16×16 overlapping block of video frame. These blocks are subtracted simultaneously and thresholded to get a clean feature representation which is compact and defines hand boundaries. Finally, the multi class support vector machine classifies gestures into signs.

Table-1. Database of video words from different human samples.

Samples	Frame#31	Frame#49	Frame#68	Word meaning
1				GO
2				COW
3				UPWARDS
4				SMALL
5				CROW

2. LITERATURE

Sign language recognition (SLR) is categorically separated into two groups: Discrete Sign representation and continuous sign representation. There are numerous approaches proposed by researchers in both modalities but a computerized system is still a distant reality. Each representation is further classified as signer-dependent and signer independent systems. Signer dependent systems are sensitive to the human signer in the video frame whereas signer independent systems are immune to human signer. Further two more approaches are prevalent in SLR:(i) glove based and (ii) vision based. Both approaches are exclusively modelled over the decades. In glove based the system used electromechanical equipment to collect data. In recent times radio frequency gloves are becoming popular with researchers. Vision based systems are complicated on the system requirement but offer ease to the signer as he must perform signs with bare hands. The classification models largely used by research community are hidden Markova models (HMM), fuzzy inference engines (FIS) and Artificial Neural Networks (ANN's).

Glove based SLR's are popularly known as cyber gloves [1] [2]. In [3] Wei *et at*., component based SLR with extendable vocabulary by using a cyber glove made of surface electromyographic (sEMG) sensors, accelerometers (ACC), and gyroscopes (GYRO). Modelling five components of sign language such as hand shape, axis, orientation, rotation and trajectory. Five subjects participated in the study with 110 signs from Chinese sign language (CSL) were classified using code matching method with an average recognition rate of 86%. Francesco Camastra *et al,* classified around 3900 hand gestures using data glove using linear vector quantization (LVQ) [4] producing a recognition rate of 99%.

Dong *et al*, classified with 90% accuracy the signs of American sign language (ASL) alphabets using a depth sensor along with a video input [5]. The main drawback with data gloves is that they are cumbersome and uncomfortable to be worn by the user [6]. Apart from RF cyber gloves, colored gloves were used in [7] to classify gestures from German sign language with 52 signs and a colour video camera sensor. Recognition rate of around 80% is reached by using K-Means clustering algorithm.

Image and Video based sign language is extensively researched in the past few years. Hand gesture image classification was extensively researched by many researchers using colour, texture and shape based features. Huong *et al*, used principle component analysis (PCA) for sparse representation of gestures of Vietnamese sign language (VSL) with an accuracy of 90% [8]. Indian sign language hand gestures are classified accurately to 91% by combining shape and texture features [9].

P.V.V. Kishore *et al,* developed a discrete video based sign language recognition system for 80 signs of Indian sign language [10] [11]. Sobel edge and morphological operators were used for segmentation and DCT for feature vector representation. Fuzzy inference engine classifies the discrete signs with an accuracy of 91%. The problem is illumination dependent model with simple backgrounds. They also proposed a background independent model using active contours for segmentation. But the damages are blurring and regularization of level set model used for segmentation which changed from frame to frame. Shape priors introduced from previous frames helped to reduce the segmentation error. Finally, they achieved a recognition rate of 93.5% for 50 videos of non-constant backgrounds with artificial neural networks (ANN's) [12].

P.V.V. Kishore *et al*, in their latest work introduced another parameter such as hand and head position tracking with optical flow to increase the accuracy of the algorithm [13]. But all these models are unable to work under occlusions from one hand or head during video capture.

Ruiduo Yang *et al*, proposed a method to handle movement Epenthesis and Hand Segmentation Ambiguities in continuous American sign language using dynamic time wrapping on video sequences [14]. M. Mohandes *et al*, used both vision and sensor model for recognizing gestures of Arabic sign language (Arsl) [15] [16].

A sign language recognizer much follows the dynamics of human body which is 3 dimensional whereas the cyber glove used 1D data and image and video sensors create 2D data. The low recognition rate of these models is due to the missing data component. Hence in the past couple of years' research has moved to developing 3D models and 3D avatar based sign language recognition systems [17] [18]. Microsoft kinetic is a RGB – D sensor which is being effectively applied to the field of motion capture and sign language recognition [19].

The application of computer vision for sign language recognition is a complex phenomenon. In this work, we use discrete set of videos of Indian sign language from the database created in [10] - [13]. The video is filtered and features are extracted using the proposed Local Cosine Feature (LCF) which comes from 2D DCT [20] and local neighbourhood operations on pixel blocks. Feature arrangement to create a sizeable feature matrix representing each sign. For faster operation, a distance classifier is proposed. Three distance classifiers are tested during experimentation. They are Euclidian (ED), Normalized Euclidian (NED) and squared Mahalanobis distance (SMD).

3. SYSTEM ARCHITECTURE - METHODOLOGY

The following Figure-1 shows the system architecture proposed in this work. A video is captured under controlled conditions with a 2MP mobile camera. Let this 2D video be represented as a 2D frame $\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y}) \rightarrow \square^{2}\ \forall\ (\boldsymbol{x},\boldsymbol{y}) \rightarrow \square^{+}$. For video the frame $\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y})$ changes with time, which is fixed at 30 frames per second. The size of the video is 640×480×3. It is resized to 256×256×3 for operational performance.

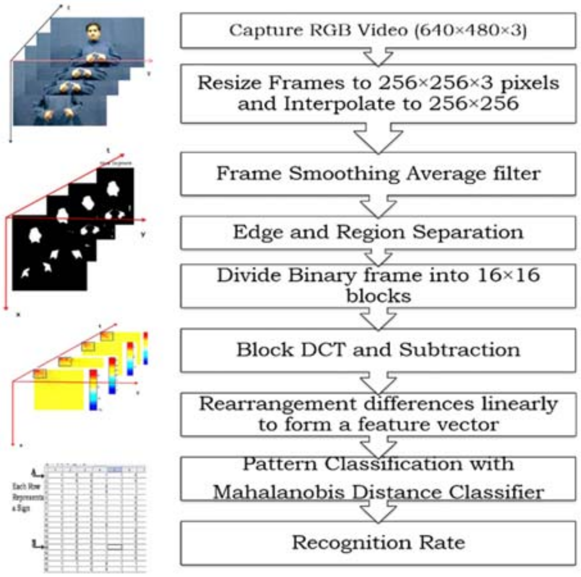
Capture RGB Video (640×480×3)
Resize Frames to 256×256×3 pixels and Interpolate to 256×256
Frame Smoothing Average filter
Edge and Region Separation
Divide Binary frame into 16×16 blocks
Block DCT and Subtraction
Rearrangement differences linearly to form a feature vector
Pattern Classification with Mahalanobis Distance Classifier
Recognition Rate
Each Row Represents a Sign

Figure-1. Proposed SLR system architecture.

Each video frame is smoothed to remove unwanted pixel variations with in the boundary regions. The RGB frames are interpolated to produce a gray tone frame. Then the frames are binarized to a dynamic range of [0, 1]. This separates the foreground hands and head portions from background.

The frames cannot be classified in higher two dimension. Hence the 2D space must be reduced to 1D subspace that represents the 2D frame contents. This task is called feature extraction. To reduce the size of the feature vector in the past many researchers have exported the spatial domain contents to frequency domain. In the previous works researchers proposed discrete cosine transform (2D DCT) to do the job.

Features are unique representation of objects in this world. Feature is a set of measured quantities in a 1D space represented as $\boldsymbol{F}^{V}(\boldsymbol{x}) = \{\boldsymbol{f}(\boldsymbol{x}) \mid \boldsymbol{x} \subseteq \square\}$, where $\boldsymbol{f}(\boldsymbol{x})$ can be any transformation or optimization model on vector $\boldsymbol{x}$. Top priority in this work is sped of execution of the proposed algorithm. Hence $\boldsymbol{f}(\boldsymbol{x})$ is considered as Discrete Cosine Transform (DCT) [18] along with Principle Component Analysis (PCA) [19]. The 2D DCT of hand contour $\boldsymbol{H}^{C}(\boldsymbol{x})$ and head contour $\bar{\boldsymbol{H}}^{C}(\boldsymbol{x})$ is computed as

$$\boldsymbol{F}_{uv}^{V} = \frac{1}{4}\boldsymbol{C}^{u}\boldsymbol{C}^{v}\sum_{x=1}^{N}\sum_{y=1}^{N}\boldsymbol{H}^{C}(\boldsymbol{x})\cos\left(\boldsymbol{u}\pi\frac{2\boldsymbol{x}+1}{2\boldsymbol{N}}\right)\cos\left(\boldsymbol{v}\pi\frac{2\boldsymbol{y}+1}{2\boldsymbol{N}}\right) \quad (1)$$

Where $\boldsymbol{C}^{u} = \boldsymbol{C}^{v} = \frac{1}{\sqrt{2}}\forall(\boldsymbol{u.v}) = 0$ and 1 elsewhere.

Similar expression with $\bar{\boldsymbol{H}}^{C}(\boldsymbol{x})$ calculated 2D DCT of head contour as $\bar{\boldsymbol{F}}_{uv}^{V}$. Figure-2 shows a colour coded representation of hand DCT features for the frame in a video sequence. The head does not change much in any of the frames captured and hence head contour DCT remains fairly constant throughout the video sequence.

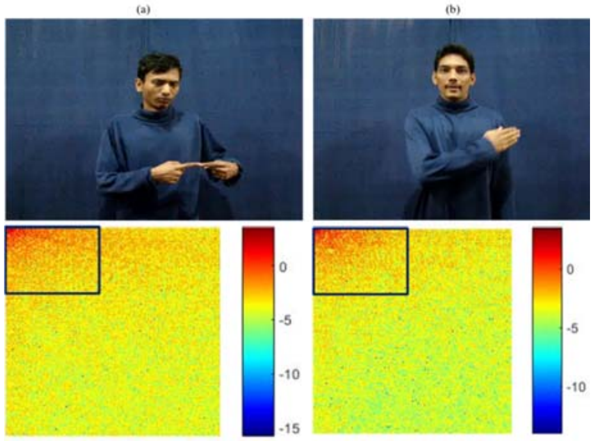

Figure-2. 2D DCT of frames (a) From sign video away and (b) from sign video GO.

Figure-2 shows both signs are different but the DCT matrix is closely packed. The variational distance between the two is quite close and this reflects in the classification stage. To ease the burden on classifier we propose a new feature vector modelled to find the variations in DCT on blocks. This projects the local DCT patterns of hands and head uniquely. The proposed method is as follows.

The binary segmented head and hand portion is divided into blocks of equal sizes. In a neighbourhood of 8 pixels in each block similar pixels are searched. The process repeats itself till the all the pixels are searched in the image. The results of such searching will separate our head and hand portions as shown in Figure-3.

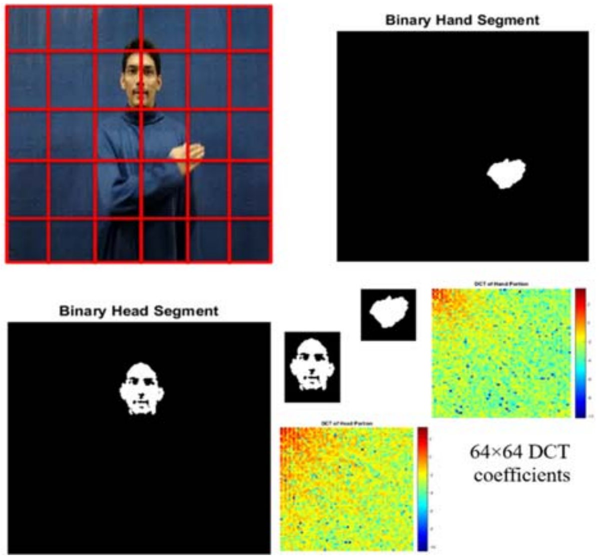

Figure-3. Hand and head separation for feature vector calculation.

Comparing Figures 2 and 3 for DCT coefficients we get coefficients that exactly model head and hand portions as in Figure-3. Reduced feature vector size is another advantage of our proposed method. Close observation of DCT coefficients in Figure-3 reveals some interesting facts about the feature vector. Most of the coefficients energy is concentrated in upper half of the matrix. Previously researchers used to choose only to the 50 rows and columns for representing the feature vector. Figure-4 shows the method followed in LZW coding for image compression.

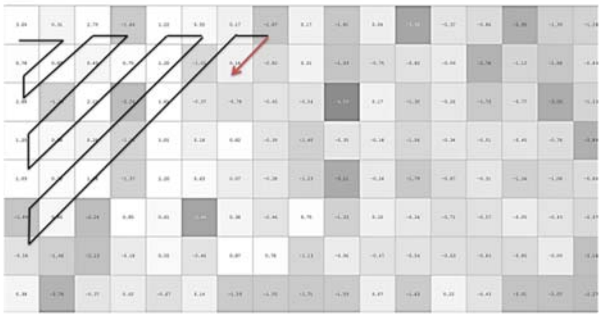

Figure-4. Feature selection process form DCT coefficient matrix.

The first 50×50 matrix of values possess maximum amount of energy in a frame. But this DCT matrix for every frame consisting of 2500 values representing a sign will cost program execution time. PCA treatment of the matrix $\boldsymbol{F}_{uv}^{V}$, which retains only the unique components of the matrix $\boldsymbol{F}_{uv}^{V}$. The final $\boldsymbol{F}_{uv}^{V}$ is represented as $\boldsymbol{F}_{fn}^{V}$, where $\boldsymbol{fn}$ gives frame number. PCA reduces the feature vector per frame to 50 sample values per frame. Each 50 sample Eigen vector from PCA uniquely represents DCT energy of the hand shape in each frame. This process does not effective in most of the cases where hand and head are close to each other.

In Figure-4 the white areas represent positive energy and dark areas negative energy. For a unique feature vector $\boldsymbol{F}_{uv}^{V}$, it should contain the more number of white portions and less number of black boxes. The selection process is indicated by the tracing line on top of the DCT coefficients of hand in Figure-3 puts an exact representation to the hand in the video frame.

The feature sign matrix $\boldsymbol{F}_{fn}^{V}$ inputs a classifier. Since speed is the prime constraint during mobile implementation, it will be reasonable to use minimum distance classifier (MDC). This is one simple classifier that does not require prior training. Mahalanobis distance is the metric that will assign class variables to different sign classes. Mahalanobis distance [24] is chosen for SLR classification on smart phones over Euclidian distance, the former includes inter sample covariance's in different directions during distance calculation. The Mahalanobis distance equals Euclidian distance for uncorrelated data with inter class variance of unity. The squared Mahalanobis distance $\boldsymbol{D}_{M}^{2}$ is given as

$$\boldsymbol{D}_{M}^{2} = \left(\boldsymbol{F}_{fn}^{V} - \boldsymbol{S}_{C}\right)^{T} \Sigma_{C}^{-1} \left(\boldsymbol{F}_{fn}^{V} - \boldsymbol{S}_{C}\right) \quad (2)$$

Where $\boldsymbol{S_C}$ is mean vector of each sign class defined by $\boldsymbol{F_{fn}^{v}}$. $\sum_C$ is the inter class covariance matrix and $\sum_C^{-1}$ is its inverse matrix. Where $\boldsymbol{T}$ is transposition. But faulty distances are measured if the inter class variance is very large. In sign language videos hand shape variations within a particular sign class are very small making Mahalanobis distance ideal for sign language classification.

4. RESULTS AND DISCUSSIONS

A mobile camera having 2M pixel resolution is used to capture video signs. The only constraint imposed on experimentation is the use of fixed contrasting background to compliment signer's hands and face portions. Lighting condition is 17-19 luminance for all the samples captured. The results are put in two sections: quantitative and qualitative. Quantitative analysis is at the image level and qualitative analysis describes the constraint handling compared to previous work in [10]. Each video is captured for a time stamp of 3 to 5 seconds depending on the movements in the sign at 30fps.

4.1. Quantitative analysis

Classification of the words is tested with Euclidian, Normalized Euclidian and Mahalanobis distance functions. Few frames of the video sequence are in Figure-5.

The words used in this paper are - foolish, fat, young, duck, peacock, camel, apart, away, big, small, feather, flat, funny, go, nest, lift, large, balance, boring and close. 10 signers participated in the experiment. A total of 20 different signs of Indian sign language are captured with 10 different signers of different shapes and sizes thus making the system signer independent.

Filtering and adaptive thresholding with sobel gradient produces regions of signer's hands and head segments. Figure-6 shows the results of the segmentation process on a few frames. Row (a) has original RGB captured video frames. Row (b) has average filtered, soble gradiented and region filled outputs of the frames in row (a). The last row contains morphological subtracted outputs of the frames in row (b)

The energy of the hand and head contours gives features for sign classification. 2D DCT calculates energy of the hand and head contours. DCT is uses orthogonal basis functions that represent the signal energy with minimum number of frequency domain samples that can effectively use to represent the entire hand and head curvatures. As shown in figure 4, only 150 samples of the DCT matrix were extracted representing each hand sign in each frame. These 150 samples out of 65536 samples are enough to reproduce the original contour using inverse DCT. This hypothesis is tested for each frame and a decision was made to consider only 150 samples for sign representation. With 150 value feature matrix per frame and an average number of frames per video at 120 frames, the feature matrix for the considered each sign in the stack is 150×120 matrix. Similar with head samples. But due to limited head variations, the head feature matrix computed in first few frames is compared and an average is standardised for all frames in that video sequence. When no hand is detected in the frame it is considered as 'No Sign'. These frames are detected as their feature matrix is having only head contour energy samples. The training vector contains a few head only sample values for such 'No Sign' detection. Three classifier are compared to test the execution speeds matching that of smart phone execution. Euclidian distance, Normalized Euclidian distance and Mahalanobis distance classifies the feature matrix as individual signs. The distance classifiers are basic set of classification models that require no training methods at fast execution speeds. The next section analyses the classifiers performance based on word recognition rate (WRR).

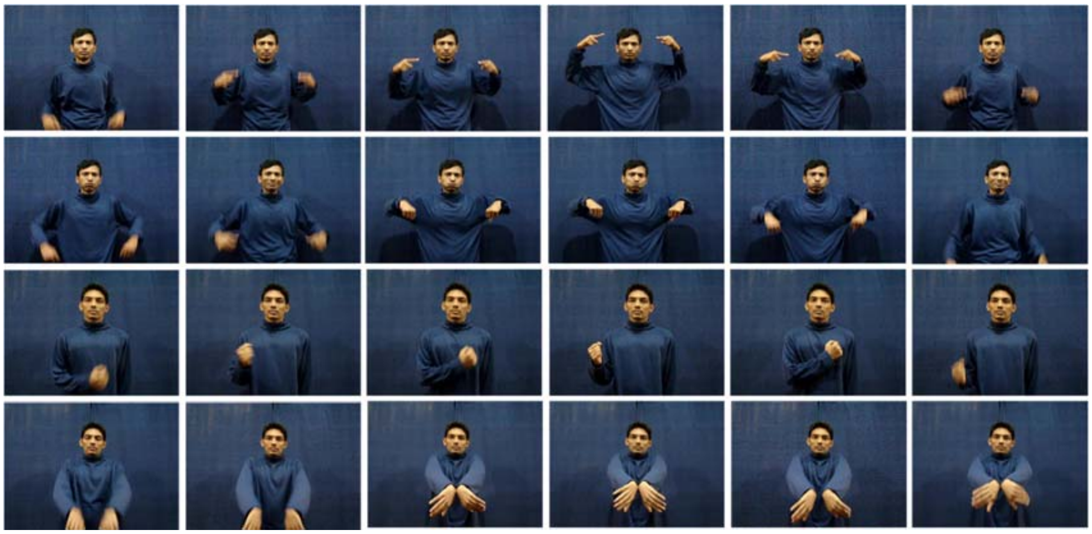

Figure-5. Row 1: sign "FOOLISH", Row 2: "FAT", Row 3: "YOUNG", Row 4: "DUCK".

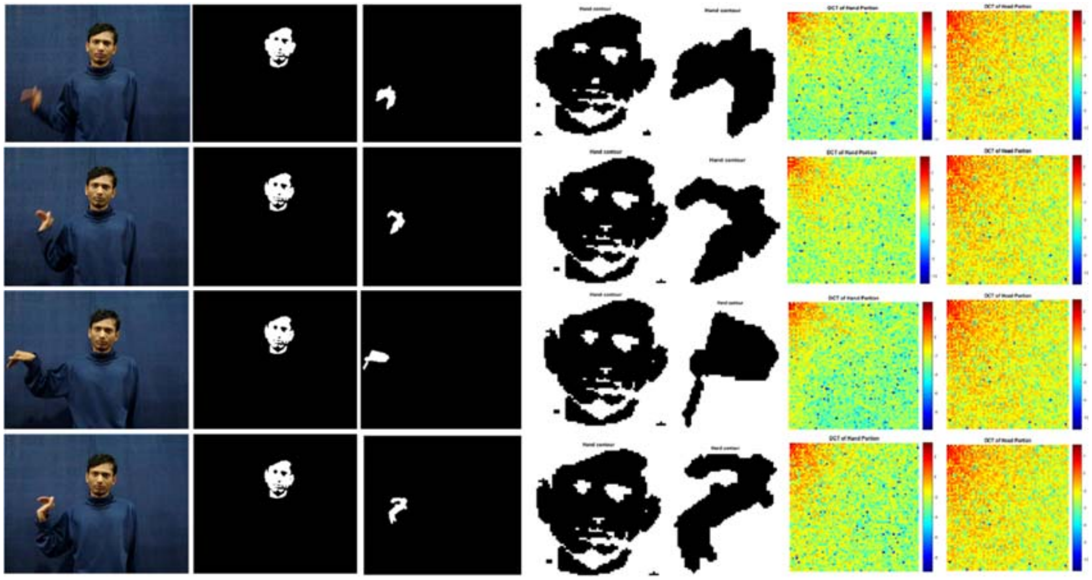

Figure-6. Row 1: frame 16 outputs and row 2: frame 28, row 3: frame 46, row 4: frame 67.

4.2. Qualitative analysis

Word Recognition Rate gives the ratio of correct classification to total number of samples used for classification. The expression for WRR

$$R^{R\%} = \frac{\textit{Correct Classifications}}{\textit{Total Signs in a Video}} \times 100\,. \tag{3}$$

Feature matrix has a size of 150×120 per sign in the video sequence. The uniqueness of a feature vector depends on its ability to represent the signs correctly. To check the features of frames in signs ‘camel’ and ‘upwards’ are plotted in Figure-7.

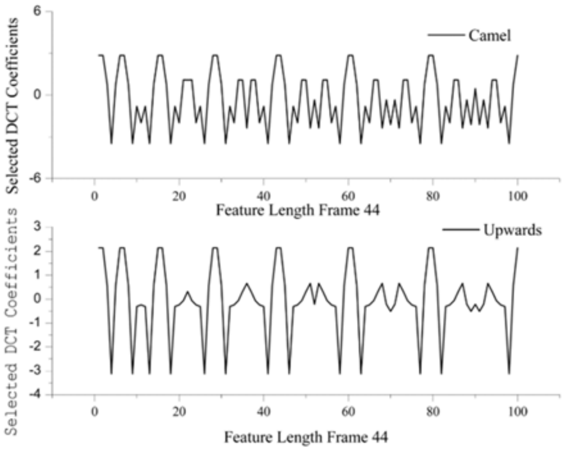

Figure-7. Feature separation between sign frames from videos of ‘camel’ and ‘upwards’.

We have conducted several experiments with ISL recognizer proposed in this work. The first experiment is evaluating the classifier by using the same data sample for training and testing. Each word consists of 18000 samples per sign per word per video. Out training and testing sample in this case is 18000×20 for a 20-word database. The classification performance of the three classifiers in catalogued in Table-2.

Table-2. Classifier performances same train test vector.

Signs	Euclidian distance classifer	Normalized euclidian distance	Mahalanobis distance classifier
Fat	80	70	90
Young	70	70	90
Duck	80	80	80
Peacock	60	60	80
Camel	90	90	100
Apart	90	90	100
Away	90	90	100
Big	90	90	100
Small	90	90	100
Feather	90	90	100
Flat	50	50	80
Funny	90	80	100
Go	60	50	80
Nest	70	70	80
Lift	50	50	90
Large	50	50	90
Foolish	60	50	80
Average WRR	**74.11**	**71.76**	**90.58**

The average classification rate with same training feature for testing individual words is around 90.58% with Mahalanobis distance. The low scores recorded by Euclidian distance (74.11%) and normalized Euclidian Distance (71.76%) compared to Mahalanobis is the inter class variance considerations as in eq'n 2. Test repetition frequency is 10 per sign.

To find the average WRR for all the different signer video samples and to test the signer independent functionality of the system we conduct our next experiment. In the next experiment, we us a feature vector from a different signer in the same order as the training data.

Table-3. Classifiers performance with different training and testing sample in same order.

Signs	Euclidian distance classifer	Normalized euclidian distance	Mahalanobis distance classifier
Fat	70	60	80
Young	60	60	80
Duck	70	70	80
Peacock	50	40	80
Camel	80	80	90
Apart	80	80	100
Away	80	80	100
Big	80	80	100
Small	80	80	90
Feather	80	80	90
Flat	40	40	80
Funny	60	80	90
Go	50	40	80
Nest	60	60	80
Lift	40	40	80
Large	40	40	80
Foolish	50	40	80
Average WRR	**62.94**	**61.76**	**85.88**

The Table-3 shows a decrease in WRR from the previous test but the values are close to be considered as good classification. Further, Table-4 shows the WRR if the training data is in different order compared to training set.

Table-4. Classifiers performance with different training and testing sample in different order.

Signs	Euclidian distance classifer	Normalized euclidian distance	Mahalanobis distance classifier
Fat	70	60	80
Young	60	60	80
Duck	70	70	80
Peacock	50	40	80
Camel	80	80	90
Apart	80	80	100
Away	80	80	100
Big	80	80	100
Small	80	80	90
Feather	80	80	90
Flat	40	40	80
Funny	60	80	90
Go	50	40	80
Nest	50	50	70
Lift	40	40	80
Large	40	40	80
Foolish	50	40	80
Average WRR	**60.94**	**60.76**	**83.22**

There is only a slight variation in WRR score and classifiers are robust to inter class data variations. Only one sign "Nest" failed slightly due to the complexity of the sign as both hands are interlocked on to one another. This generates occlusions and thus the system has no mechanism to handle such issues.

Head features do not change from frame to frame in most of the signs and this is understood from the plots in Figure-8.

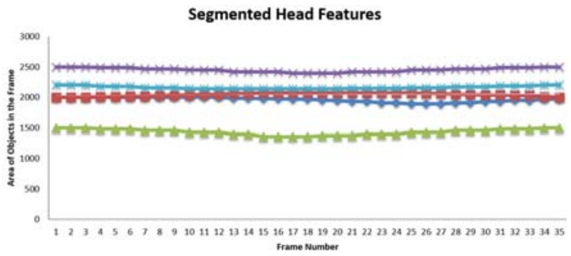

Figure-8. Head changes in consecutive frames for signs shown in Table-1.

To validate the proposed method over the existing one in [11] which uses DCT features on ISL against the proposed local cosine variations as features for ISL is shown in Table-5. By using local cosine features the proposed method increased the recognition rate by 10% over the existing work in [11]. After exclusive experimentation for a number of times the average recognition rate stands at 90.44%.

Table-5. Comparison of proposed method with that in ref [11].

Sign	Correctly recognized signs	WRR (%) [11]	Correctly recognized signs	WRR proposed method
Fat	10	100	10	100
Young	9	90	10	100
Duck	9	90	10	100
Peacock	8	80	9	90
Camel	10	100	10	100
Apart	5	50	7	70
Away	7	70	8	80
Big	7	70	8	80
Small	8	80	9	90
Feather	7	70	8	80
Flat	10	100	10	100
Funny	8	80	9	90
Go	9	90	9	90
Nest	5	50	7	70
Lift	9	90	10	100
Large	9	90	10	100
Foolish	9	90	10	100
TOTAL	**139/170**	**81.76**	**154/170**	**90.58**

The confusion matrix for the test data is shown in Figure-9. Confusion matrix plots the total number of miss-classifications and the reason for that miss-classification can be interpreted form the video data.

Signs	Fat	Young	Duck	Peacock	Camel	Apart	Away	Big	Small	Feather	Flat	Funny	Go	Nest	Lift	Large	Foolish	Mis-Class
Fat	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Young	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Duck	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Peacock	0	0	1	9	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Camel	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
Apart	0	0	0	0	0	7	2	0	0	0	0	0	1	0	0	0	0	3
Away	0	0	0	0	0	1	8	0	0	0	0	0	1	0	0	0	0	2
Big	0	0	0	0	0	0	0	8	1	0	0	0	0	0	0	1	0	2
Small	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	1	0	1
Feather	0	0	1	1	0	0	0	0	0	8	0	0	0	0	0	0	0	2
Flat	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0
Funny	0	0	0	0	0	0	0	0	0	0	0	9	1	0	0	0	0	1
Go	0	0	0	0	0	0	0	0	0	0	0	1	9	0	0	0	0	1
Nest	0	0	0	0	0	1	1	0	0	1	0	0	0	7	0	0	0	3
Lift	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0
Large	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0
Foolish	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0

Figure-9. Confusion matrix for test data.

5. CONCLUSIONS

In this work, we have shown a way to recognize gestures of Indian Sign Language through videos captured from mobile camera. Using Mahalanobis distance classifier for signer independent recognition was averaged at 90.44% compared to other type of distance measures such Euclidian and Normalized Euclidian distance classifiers. This achievement is important for ISL as there are only few research works to compare. We also showed that there is a significant improvement in misclassification by using LCF instead of DCT on the entire video frames. This was a significant improvement in WRR. The work was also validated by comparing with some existing works on ISL. There are still many issues related to ISL recognition such as two hands, shoulder detection, face expressions and occlusion management. There are many feature extraction models still to be explored along with some machine learning algorithms.

REFERENCES

[1] Cheng J.; Chen X.; Liu A.; Peng H. 2015. A Novel Phonology-and Radical-Coded Chinese Sign Language Recognition Framework Using Accelerometer and Surface Electromyography Sensors. Sensors. 15, 23303-23324.

[2] Wang Hanjie, Xiujuan Chai and Xilin Chen. 2016. Sparse Observation (SO) Alignment for Sign Language Recognition. Neurocomputing. 175: 674-685.

[3] Wei S., Chen X., Yang X., Cao S. & Zhang X. 2016. A Component-Based Vocabulary-Extensible Sign Language Gesture Recognition Framework. Sensors. 16(4): 556.

[4] Camastra F.; De F.D. 2012. LVQ-based hand gesture recognition using a data glove. In Proceedings of the 22nd Italian Workshop on Neural Networks, 2012, Vietri sul Mare, Salerno, Italy. pp. 159-168.

[5] Dong C.; Leu M.; Yin Z. 2015. American Sign Language Alphabet Recognition Using Microsoft Kinect. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA. pp. 44-52.

[6] Sun C.; Zhang T.; Bao B.K.; Xu C.; Mei T. 2013. Discriminative exemplar coding for sign language recognition with Kinect. IEEE Trans. Cybern. 43, 1418-1428.

[7] H. Hienz, B. Bauer, K.F. Krais. 1999. HMM-based continuous sign language recognition using stochastic grammar, in: Proceedings of GW'99, LNAI 1739, pp. 185-196.

[8] Huong, Thao Nguyen Thi, Tien Vu Huu, and Thanh Le Xuan. 2015. Static hand gesture recognition for vietnamese sign language (VSL) using principle components analysis. In 2015 International

Conference on Communications, Management and Telecommunications (ComManTel). pp. 138-141. IEEE.

[9] Kishore P. V. V., S. R. C. Kishore and M. V. D. Prasad. 2013. Conglomeration of hand shapes and texture information for recognizing gestures of Indian sign language using feed forward neural networks. International Journal of engineering and Technology (IJET), ISSN: 0975-4024.

[10] Kishore P. V. V. and P. Rajesh Kumar. 2012. A video based Indian sign language recognition system (INSLR) using wavelet transform and fuzzy logic. International Journal of Engineering and Technology. 4.5: 537.

[11] Kishore P. V. V., Kumar P. R., Kumar E. K. & Kishore S. R. C. 2011. Video Audio Interface for Recognizing Gestures of Indian Sign. International Journal of Image Processing (IJIP), 5(4), 479.

[12] Kishore P. V. V. and P. Rajesh Kumar. 2012. Segment, Track, Extract, Recognize and Convert Sign Language Videos to Voice/Text. International Journal of Advanced Computer Science and Applications (IJACSA) ISSN (Print)-2156 5570.

[13] Kishore P. V. V. and M. V. D. Prasad. 2015. Optical Flow Hand Tracking and Active Contour Hand Shape Features for Continuous Sign Language Recognition with Artificial Neural Networks. International Journal of Software Engineering and Its Applications. 9(12): 231-250.

[14] Ruiduo Yang, Sudeep Sarkar, and Barbara Loeding. 2010. Handling Movement Epenthesis and Hand Segmentation Ambiguities in Continuous Sign Language Recognition Using Nested Dynamic Programming. IEEE transactions on pattern analysis and machine intelligence. 32(3): 462-478.

[15] Mohandes, Mohamed, Mohamed Deriche and Junzhao Liu. 2014. Image-based and sensor-based approaches to Arabic sign language recognition. IEEE Transactions on Human-Machine Systems. 44(4): 551-557.

[16] A. S. Elons, M. Abull-Ela, and M. F. Tolba. 2013. Pulse-coupled neural network feature generation model for Arabic sign language recognition. IET Image Process. 7(9): 829-836.

[17] Wu Xingyu, Xia Mao, Lijiang Chen and Yuli Xue. 2015. Trajectory-based view-invariant hand gesture recognition by fusing shape and orientation. IET Computer Vision. 9(6): 797-805.

[18] Huang Jie, Wengang Zhou, Houqiang Li and Weiping Li. 2015. Sign language recognition using 3D convolutional neural networks. In 2015 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6. IEEE.

[19] Amin, Omar, Hazem Said, Ahmed Samy and Hoda K. Mohammed. 2015. HMM based automatic Arabic sign language translator using Kinect. In Computer Engineering and Systems (ICCES), 2015 Tenth International Conference on, pp. 389-392. IEEE.

[20] Burger Wilhelm and Mark J. Burge. 2016. The Discrete Cosine Transform (DCT). In Digital Image Processing, pp. 503-511. Springer London.

[21] Dahmouni, Abdellatif, Nabile Aharrane, Karim El Moutaouakil and Khalid Satori. 2016. Face Recognition Using Local Binary Probabilistic Pattern (LBPP) and 2D-DCT Frequency Decomposition. In 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV), pp. 73-77. IEEE.

[22] J. Mei, M. Liu, Y. F. Wang and H. Gao. 2016. Learning a Mahalanobis Distance-Based Dynamic Time Warping Measure for Multivariate Time Series Classification. in IEEE Transactions on Cybernetics. 46(6): 1363-1374.

[23] N. Kim. 2010. Euclidian distance minimization of probability density functions for blind equalization. in Journal of Communications and Networks. 12(5): 399-405.

[24] J. Mei, M. Liu, Y. F. Wang and H. Gao. 2016. Learning a Mahalanobis Distance-Based Dynamic Time Warping Measure for Multivariate Time Series Classification. in IEEE Transactions on Cybernetics. 46(6): 1363-1374.

[25] Chang Kang Tsung. 2009. Computation for Bilinear Interpolation. Introduction to Geographic Information Systems. 5th ed. New York, NY: McGrawHill, Print.

[26] Gonzalez R. C. & Woods R. E. 2002. Digital image processing (2nd ed.). NJ: Prentice Hall.

[27] Fu K. and Mui J. 1981. A survey on image segmentation. Pattern Recognition. 13(1): 3-16.

[28] Chen S.; Lin W. and Chen C. 1992. Split-and-merge image segmentation based on localized feature analysis and statistical tests. CVGIP: Graph. Models Image Process. 53(5): 457-475.

[29] Haralick R. and Shapiro L. 1985. Survey, image segmentation techniques. Comput. Vision Graphics Image Process. 29(1): 100-132.

[30] Pal N. and Pal S. 1993. A review on image segmentation techniques. Pattern Recognition. 26(9): 1277-1299.

[31] Liu Cai. 2004. A kind of advanced Sobel image edge detection algorithm. Guizhou Industrial College Transaction (Natural Science Edition). 33(5): 77-79.

[32] J. Serra. 1982. Image Analysis and Mathematical Morphology. Academic Press, London.

[33] E. R. Dougherty, J. Astola. 1994. An Introduction to Nonlinear Image Processing. SPIE, Bellingham, WA.