VOL. 13, NO. 4, FEBRUARY 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

SIGN LANGUAGE RECOGNITION WITH MULTI FEATURE FUSION AND ADABOOST CLASSIFIER

P. Praveen Kumar, P. V. G. D. Prasad Reddy and P. Srinivasa Rao
Department of Computer Science and Systems Engineering, College of Engineering Andhra University, Visakhapatnam, India
E-Mail: pkpinjala.auce@gmail.com

ABSTRACT

Extracting and recognizing complex human movements from video sequences is a challenging task. In this paper a complicated problem from the class is approached using Indian sign language videos. A new segmentation model is developed using discrete wavelet transform and local binary pattern (LBP) features for segmentation. A 2D point cloud is created from the local sign shape changes in subsequent video frames. The classifier is fed with 2 types of features calculated from Global Haar features and Local LBP features. We also explore multiple feature fusion models after segmentation for improving the classification process with state of the art features such as HOG, SIFT and SURF. The extracted features input the Adaboost multi class classifier with labels forming the corresponding words. We test the classifier on Indian sign language video dataset prepared in controlled environments. The algorithms were tested for accuracy and correctness in identifying the signs.

Keywords: Indian sign language identification, adaboost classifier, multi feature fusion, discrete wavelet transform, local binary patterns.

1. INTRODUCTION

Automatic sign language recognition is a complicated problem for computer vision scientists, which involves mining and categorizing spatial patterns of human poses in videos. Sign language created from human action is defined as a temporal variation of human body in a video sequence, which is characterized by moving hands with respect to body, face, head including hand shapes. Automation encompasses mining the video sequences with computer algorithms for identifying similarities between actions in the unknown query dataset with that of the known dataset. Last decade has seen a jump in online video creation and the need for algorithms that can search within the video sequence for a specific human pose or object of interest. The problem is to extract, identify a human pose and classify into labels based on trained human signature action models [1] - [3]. The objective of this work is to extract the signature of Indian sign language poses from videos given a specific sign as input. However, the constraints are video resolution, frame rate, background lighting, scene change rate and blurring to name a few. The analysis on video content is a complicated process as the most of the users end up with constraints which act as a hindrance in automation of video object segmentation and classification. Sign language video sequences are having far many constraints for smooth extraction of sign signatures. Automatic sign extraction is complicated due to complex hand poses and body actions performed at different speeds depending on the signer. Figure 1 shows a set of lab captured Indian sign videos for training and testing the proposed algorithm.

Figure-1. Sign language datasets used in this work captured from various sensors at different object distances and background lighting variations.

Sign language is a visual mode of communication between two hearing impaired or hard hearing people. The communication foundations are based on finger shapes, hand shapes, hand movements in space with respect to body, hand orientations and facial expressions. The humans are trained exclusively to handle such huge amounts of information for years. For machine translation, the problem transforms into a 2D natural language processing problem. Many 1D/2D/3D models are proposed in literature with little success to bring the model close to real time implementation [4]- [8].

Extracting these complex movements from videos and classification requires a complex set of algorithms working in sequence. We propose to use silhouette detection and background elimination, human object extraction, local texture with shape reference model and 2D point cloud to represent the signers pose. Global and Local features are calculated that represents the exact shape of the signers' hand in the video sequence. For recognition, a multiclass multilabel Adaboost algorithm is proposed to classify query sign video based on the Indian sign language dataset.

The rest of the paper is organized into literature survey on the proposed techniques, theoretical background

on the proposed models and experimental results. The proposed model is compared with different state of the art features such as Histogram of Oriented features (HOG), scale invariant feature transform (SIFT) and speed up robust features (SURF) with classifiers such as SVM classifier already proposed by us in the previous works.

2. LITERATURE

Sign language recognition (SLR) has transformed with technology upgradation from 1D, 2D to 3D models in the last 2 decades. In 1D, SLR is based on 1D signals acquired from a hand gloves [8] and classified using signal processing methods [9]. In the recent times researchers started using leap motion sensor [10] to extract 1D signals of finger movements and estimate the related gestures of sign language using Hidden Markov Models.

The faster 1D models produce good recognition rates when the emphasis is only on hands. But sign language involves head, torso and face expressions along with hand movements and shapes [11]. 2D video data of signs produces relatively more information compared to 1D data gloves. From 2D capture, one can explore all the elements of a visual language with a constraint on speed and classification accuracy. Again, for 2D SLR HMM is most widely researched classifier with continuous and discrete versions of sign language [12]. More research related material on 2D models and the corresponding research challenges can be found in [13]- [15]. The other challenge for researchers lies in converting the detected signs into meaningful sentences [11]. The challenging problems in 2D SLR are hand tracking, occlusions on hands and face, background lighting, changing signer backgrounds and camera sensor dynamics.

Unlike America and Europe, India does not have a standard sign language with necessary variations. However, recently, Ramakrishna Mission Vivekananda University, Coimbatore developed an ISL dictionary with approximately 2037 signs [16] currently available. SLR systems are classified into two broad categories: sensor glove-based [17] and vision-based systems [18, 19].

Starner proposed a real-time ASL recognition system with a wearable computer-based video [18], which uses a hidden Markov model (HMM) for continuous recognition of ASL. In this system, signs are modeled with four states of HMMs with high recognition accuracies. However, this system is not signer independent.

Bhuyan *et al*. [20] used hand shapes and hand trajectories to recognize static and dynamic hand signs from ISL. They used the object-based video abstraction technique for segmenting the frames into video object planes by considering the hand as a video object. Their experimental results reveal that this system can recognize and classify static and dynamic gestures as well as sentences with superior consistency.

Akmelia *et al.* [21] proposed a real-time Malaysian sign language translation system by using the color segmentation technique, with a recognition rate of 90%. Habili *et al*. [22] proposed a hand and face segmentation technique that used color and motion cues for content-based representation of sign language video sequences.

Zhou and Chen [23] proposed a signer adaptation method, in which maximum a posteriori estimation was combined with iterative vector field smoothing to reduce the amount of data to be translated. This method achieved high recognition rates. Moreover, gesture recognition systems that employ statistical approaches [24], example-based approaches [25], and finite-state transducers [26] have exhibited higher translation rates.

In the past decade, with the advances in efficient computing and bigger parallel corpora, more efficient algorithms have been developed for training [27] and generation [28]. A considerable amount of work on Indian sign language recognition is being accomplished in [29]-[35] with machine learning algorithms on both discrete and continuous sign languages. In this work, we use the dataset from their work.

The objective is to select features that represent a sign and is easily distinguishable in closely related sign words and are computationally efficient. The attributes for a sign language recognizer chosen are global shape features using Haar wavelet [35] for hand and body shapes. However, Haar global shapes fail to characterize localized hand movements that are scuttle with respect to spatial hand movements in the video. The small hand variations are captured using 2D point cloud generated from harr wavelet and local binary pattern (LBP) [36] which attributes to local features. The chosen attributes perfectly characterize a sign in Indian sign language.

Classifying at faster rate on a huge dataset is a complicated problem. Adaboost [37] classifier is faster and efficient algorithm for large datasets [38]. Inspired from [39] and [40], the feature matrix is labelled and inputted to Adaboost classifier for training and testing. The performance indicators are recall-precision curves and execution time on mobile are recorded to check the robustness of the algorithm and feasibility to implement more efficiently.

In this paper, we propose an multi class multi label Adaboost (MCMLA) based classification problem on multidimensional feature vector. We show that this can be used to match large unconstrained sign features which are automatically extracted from video datasets. The feature representation of video objects depends on the efficiency of video segmentation algorithms. As illustrated in figure 2, the proposed Adaboost can effectively recover the query video frames from the dataset, by global shape - local shape observation model defied by discrete wavelet transform (DWT) and local binary patterns (LBP).

In summary, our MCMLA algorithm on Indian sign language videos combines the representational flexibility and trivial computations. We perform experiments on two different datasets of Indian sign language with different controlled backgrounds. The proposed method is compared with state of the art features and SVM classifier which are outperformed by a considerable margin in accuracy.

3. PROPOSED METHODOLOGY

The proposed algorithm framework is shown in Figure-2. An Indian sign language video library is created

combining black video backgrounds. Signer identification, signer extraction, global and local shape feature extraction and classifier are the modules of the system. Further feature fusion concept from [41] is also explored in this wok using two feature types, made from LBP features and Haar features. Adaboost algorithm explores the relativity between the query sign sequence and known dataset.

3.1 Signer identification

Most of the videos are poorly illuminated or fully brightened with too much background information during real time capture. Commercial video cameras have a frame rate of 30fps and the hand movements are sometimes faster and at time slower which makes the object blurry. The objective is to extract moving signer and segment it for further processing.

This helps to prevent the algorithm form constantly upgrade the background information and model the object characteristics in real time. The signer identification module is based on one of the silhouette extraction methods proposed in [42]. A significant indication in determining hand motion for extraction lies in the temporal changes in the signers' silhouette during signing. To avoid background modelling and foreground extraction models, we propose to use the following procedure.

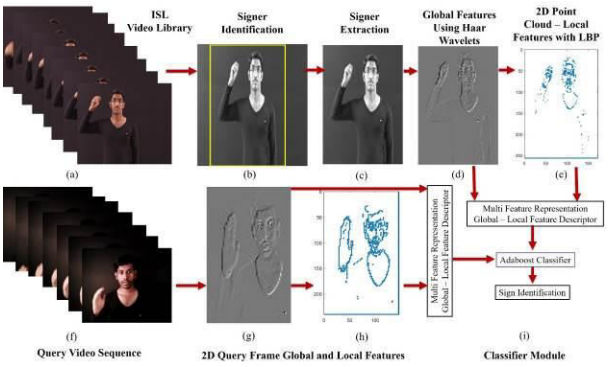

Figure-2. Flow diagram of the proposed process for Indian sign language recognition. (a) Training datasets, (b) Detected signer object, (c) Extracted, (d) DWT and LBP features (e) 2D Feature points, (f) Query dance video, (g),(h) Query feature points, (i) Multi feature extraction and classification.

The sign video sequence $V(x,y,t) \subset \mathbf{R}^+$, with $(x,y) \subset \mathbf{Z}^+$ gives pixel location and $t \subset \mathbf{Z}^+$ is the frame number. Each frame in V is having RGB planes and is of size $N \times M \times 3$. This part of the module is only for motion segmentation and object extraction; color can be discarded. RGB is converted to gray scale and contrast enhanced to improve the frame quality. The frame V^t at t is mean filtered with mask defined by $m(x,y)$ with

$$V_m^t(x,y) = V^t(x,y) \otimes m(x,y) \quad (1)$$

The size of m is updated based on the frame size $N \times M$ for faster computations, where the object area is small compared to the background area. The $\otimes$ operator is linear convolution and the averaged frame is of same size as the input frame. The next step applies a Gaussian filter of μ mean and σ variance on the input frame V^t

$$V_g^t(x,y) = V^t(x,y) \otimes g(\mu,\sigma) \quad (2)$$

The size of the Gaussian mask is determined by the input video frame. Euclidian distance metric $S^t(x,y)$ between V_m^t and V_g^t gives the saliency map of the moving pixels in the frame

$$S^t(x,y) = \left\| V_g^t(x,y) - V_m^t(x,y) \right\|_2 \quad (3)$$

The second order normed distance map is shown in figure 3 which identifies the signer's silhouette. However, to extract the signer, a mask of this silhouette is used to determine the connected components in the object. Figure-3(e) shows the silhouette mask.

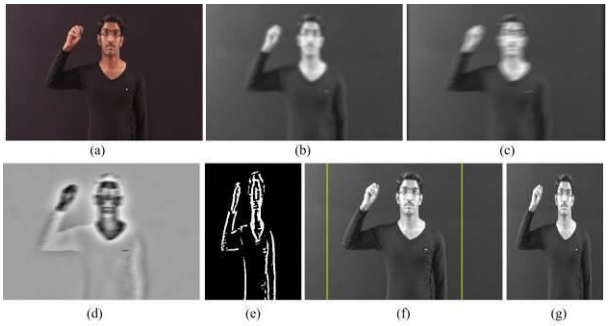

Figure-3. Signer extraction. (a) Original frame, (b) Gaussian filtered, (c) Average filtered, (d) Distance Saliency map, (e) Silhouette mask, (f) Detected signer and (g) Signer extraction.

The centroid of the mask is mapped on the frame to crop out the signer in the frame. The method is effective in all lighting conditions putting constraints on the input video frame size in selecting the masks used for mean and Gaussian filters. The boxed and extracted signer from the video sequence is shown in figure.3. (g) and (h) respectively. The extracted signer is free from background variations in the video sequence. If a portion of background still appears at this stage can be nullified during the matching phase. Applying feature extraction on the extracted signer allows for lesser computations as the background is almost eliminated and leads to good matching accuracy.

3.2 Sign feature extraction

From a signers' perspective, to identify a sign type in any sign language, hand shapes and their movements in space are the vital features. Feature extraction phase explores the methodology in extracting

these two features. There are many shape descriptors available in literature for characterizing shape features [43]. Lighting, frame inconsistency, contrast, blurring and frame size are some of the critical factors that affect feature extraction algorithms. In addition, the dancer velocity during performance instincts for a faster shape extractor.

3.2.1. Haar wavelet features - Global shape descriptor

For removing video frame noise during capture and to extract local shape information, we propose a hybrid algorithm with discrete wavelet transform (DWT) [35] and Local Binary Patterns (LBP) [36]. The objective at this stage is to represent moving dancers shape with a set of wavelet coefficients. Here we propose to use Haar wavelet at level 1. At level 1, Haar wavelet decomposes the video frame V^t into 4 sub-bands. Figure-4 shows the 4 sub-bands at 2 levels. At 1st level we have 4 sub-bands and at 2nd level have 8 sub-bands. In the 1st level, the three sub-bands represent the shape information at three different orientations: Vertical υ, Horizontal h and Diagonal d. Combining the three sub-bands and averaging the wavelet coefficients normalizes the large values.

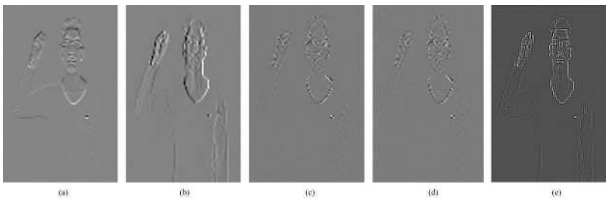

Figure-4. Harr Wavelet sub-bands: Global shape Representation, (a) Horizontal, (b) Vertical, (c) Diagonal, (d) Averaged h,v,d and (e) Reconstructed global shape matrix.

$$W_S^t = \frac{h+\upsilon+d}{3} \tag{4}$$

The averaged shape harr wavelet coefficients W_S^t along with $\{h,\upsilon,d\}$ sub-band coefficients are reconstructed to spatial domain. Figure-4 shows the reconstructed spatial domain frame producing the exact global hand and body shapes of the signer. These global shape features can be used as features for recognition. However, background noise is still a major concern at this stage which blocks the minor local hand variations. Local pixel information becomes vital in classification process of hand shapes.

3.2.2. Detailed local binary patterns - local hand shapes

Apply threshold on the reconstructed ICD video frame V_r^t as

$$T^t = \sqrt{\frac{1}{NM}\sum_{j=1}^{M}\sum_{i=1}^{N}\left(V^t(j,i)\right)^2} \tag{5}$$

The binarized video frame B^t is

$$B^t = V_r^t > T^t \tag{6}$$

To extract the nodes for the graph, local pixel patterns provide exact shape representation. LBP compares each pixel in a pre-defined neighbourhood to summarize the local structure of the image. For an image pixel $B^t(x,y) \in \Re^+$, where (x,y) gives the pixel position in the intensity image. The neighbourhoods of a pixel can vary from 3 pixels with radius $r=1$ or a neighbourhood of 12 pixels with $r=2.5$. The value of pixels using LBP code for a centre pixel (x_c,y_c) is given by

$$L_S^t = LBP(x_c,y_c) = \sum_{j=1}^{P} B^t\left(g_p - g_c\right)2^p \tag{7}$$

$$B^t(x) = \begin{cases} 1 & \forall\ x \geq 0 \\ 0 & Otherwise \end{cases} \tag{8}$$

Where g_c is binary value of centre pixel at (x_c,y_c) and g_p is binary value around the neighbourhood of g_c. The value of P gives the number pixels in the neighbourhood of g_c. The local shape descriptor L_S^t of the signers' hand pose projects maximum number of points on to 2D point cloud.

3.3 Haar - LBP PCA fused features

Haar and LBP features are enough to label the signers poses in the frames and put them under the classifier. However, fusion of features at the segmentation stage improves the classification accuracy. The fusion operator is principle component analysis (PCA). PCA of wavelet Haar features and LBP features is concatenated using the following expression

$$E_{fu} = PCA\left(W_S^t\right) \cup PCA\left(L_S^t\right) \tag{9}$$

However, 3 more state of the art features such as HOG, SIFT and SURF are proposed for testing against the features described in this work that can effectively represent shape features of the signer.

3.4 Multi features - LBP, Haar

Figure-5 shows the extracted signer represented with LBP, Haar wavelet features, Detailed LBP (DLBP) features and principle component analysis fused Haar-LBP features. Detailed LBP (DLBP) accommodates both global and local hand shape features represented in a single feature vector. Given a motion frame in a ISL video sequence V^t and successfully extracted local shape features L_S^t and transformed into a binary shape matrix B_S^t of ones and zeros using equation 5. A sparse representation of B_S^t eliminates all zeros and retains only

ones and their locations in $M_S^t(x,y,\omega)$, where x,y are shape point locations and ω is shape feature weight vector.

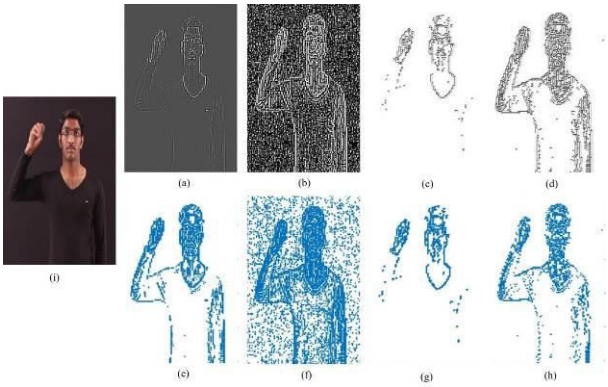

Figure-5. (a) Haar wavelet features, (b) LBP features (c) PCA fused Haar - LBP features, (d) LBP from reconstructed and thresholded wavelet coefficients (DLBP) and (e-h) sparse coded features.

Figure-5(e-h) show a sparse representation for Haar features (HF), LBPF, PCA_H_LBP and DLBP features respectively. The points on the motion object are formed by extracting the location of the pixel and its feature value determines the shape of the signer pose. From these feature point locations and values, a feature vector is constructed in this work.

3.5 Feature matrix construction

From the features, a complete feature matrix is constructed per sign frame. This feature matrix is constructed for each of the 4 features. It is a 2D feature matrix of size $256\times256\times1$ with most of the values being zeros. Sparse pooling is performed to reduce the feature set to $3\times N$ per frame, where N is number of non-zero elements in the features. The three rows of the matrix indicate pixel value and the other two rows are pixel location. We calculated, the final feature matrix as $3\times N$ per frame. These features are carefully labelled with vocal words representing the sign form in the video frame. For a 125-frame word *'GOOD'*, we have $3\times N\times125$ feature matrix. To minimize it into a 2D matrix, the value of N is chosen as the maximum in all frames. If in a frame, $N<\max(N)$, the difference locations are zero padded. The matrix is reshaped to $3N_{Max}\times125$ feature words. This feature matrix or a set of matrices are inputted to MCMLAB classifier.

3.6 Sign classifier: Adaboost multi-class multi-label

Boosting based classifications [44]- [46] finds very precise hypothesis from a set of weak hypothesis. Here hypothesis is a classification rule. The set of weak hypothesis are simple rules that generate a predictable classification. Let $\mathrm{T}=\left[(f_1,L_1),(f_2,L_2),(f_3,L_3),.....,(f_\upsilon,L_\upsilon)\right]$ be a set of training examples at an instance f_i on i^{th} frame in feature space f with labels L_i on label space L. The algorithm accepts the training samples T along with some class distribution $D=\{1,....,m\}\in\mathbf{R}$ represented as weak learners. On the input, the weak learner computes a weak hypothesis H. Generally, $H:f\rightarrow\mathbf{R}$. The interpretation for classification is based on $sign\{H(f)\}=\{+1,-1\}\rightarrow\{f_i\}$ for a binary classifier. The $|H(f_i)|$ gives prediction confidence.

The key to boosting is to use the weak learner to produce a very precise prediction rule by repeatedly addressing the weak learner on different distribution of training examples. In this work, a multiclass version of Adaboost is used having a set of strings as class labels. The problem is modelled as given T and size of final strong classifier C, The Adaboost initializes the distribution function as

$$D_1(i,l)=\frac{1}{\upsilon m}\;\forall i=1,..,\upsilon\;\&\;l=1,..,m \tag{10}$$

Where $l=|L|$. For $c=1,...,C$, we select a weak classifier $H_c:\mathrm{T}\times L\rightarrow[-1,1]$ with distribution D_c, to maximize the absolute value of

$$a_c=\sum_{i,c}D_c(i,l)L_i(l)H_c(f_i,l) \tag{11}$$

We choose the biasing value α_c as

$$\alpha_c=\frac{1}{2}\ln\left(\frac{1+a_c}{1-a_c}\right) \tag{12}$$

and update the distribution function as

$$D_{c+1}(i,l)=\frac{D_c(i,l)e^{-\alpha_c L_i[l]H_c(f_i,l)}}{N_c} \tag{13}$$

Where N_c is normalization factor to keep the distribution as probability density function. The final output strong classifier is

$$H(f,l)=sign\left(\underset{c}{sum}\left(\alpha_c H_c(f,l)\right)\right) \tag{14}$$

For the multi class problem $c_1,....,c_k$, we use the real valued 2D Look Up Table model in [47], which is defined as

$$H_{LUT}(f,L)=\sum_{i=1}^{n}\sum_{j=1}^{k}\left(2P_l^{(j)}-1\right)B_n^{j,l}(f,l) \tag{15}$$

Where, $P_l^{(j)} = P(f \in C_l \mid frame \in sign)$ and $B_n^{i,l}(f,l) = \begin{cases} 1, & f \in 1 \\ 0, & otherwise \end{cases}$ .From this weak hypothesis, through training a strong hypothesis is generated to recognize sign labels $Z_i = H(f_i)$.

4. EXPERIMENTATION AND RESULTS

A set of 4 experiments are introduced to test the strength of the proposed features from Haar, LBP, DLBP and PCA_H_LBP with Adaboost classifier. Our Indian sign language datasets consist of videos captured in controlled environment at K.L. University, cams department studio with black background. We have created 151-word continuous sign videos from 5 different signers with a total dataset of 755 sign words from Indian sign language.

Each word is labelled with its name and transitions between words is named as 'Null'. All the frames in a sign video are given the same label. The database hosts a set of features computed for each word and their transitional features in sequential order. The database sequence is arranged phonetically with the words used in this work. Only a single combination of words is chosen in this work and more random combination of words can be used in our future works.

We use percentage of recognition as a performance evaluator for validating the results. For a strong hypothesis H resulted from Adaboost training and testing for an input feature f_i on a trained distribution D with $Z_i = H(f_i)$ predicted labels. The following metrics in [49] are

$$\text{Re}\,cognition = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{|L_i \text{ I } Z_i|}{|L_i \text{ U } Z_i|} \quad (16)$$

The first experiment, 'Exp-I', uses Haar features to represent the signer. The Haar features give a global perspective on the sign in a video frame. Figure-6(b) shows Haar features on a set of video frames for the sign 'GOOD'. The Adaboost classifier is trained with Haar feature matrix of individual sign with a predefined label. The label used is a string to represent the sign meaning. The testing input is a video of continuous signs with a phonological meaning. Some of the words in the testing continuous sign input are "Hello, Good Morning, how is your day. Beautiful day. Without any worries in the world. My mother is hard working woman. My father is a software professional". Figure 7 shows the confusion matrix after testing with MCMLAB classifier on Haar features. The average recognition per word is around 78.96%.

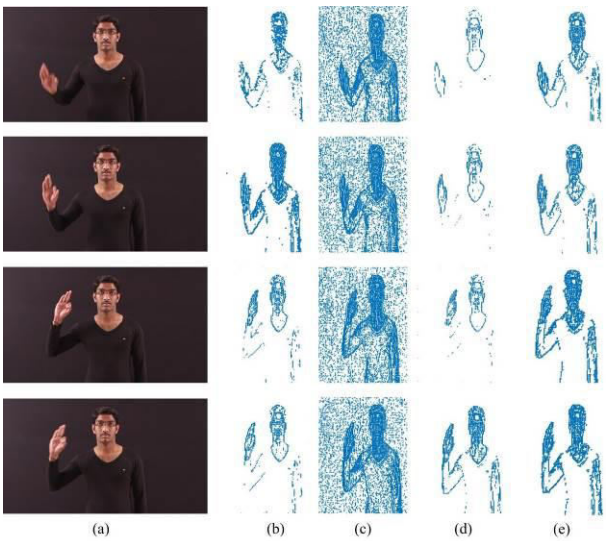

Figure-6. (a) Original video frames of sign 'GOOD', (b) Haar features, (c) LBP, (d) PCA_H_LBP and (e) DLBP features.

Figure-7 shows a set of signs used in the 151 length word sentences used as input as continuous Indian sign language. Similar looking signs have a greater percentage of matching and exhibits closeness to each other compared to differently looking signs. The database is a mixture of 5 signers with different knowledge of sign language with the most experienced signer videos are used for training the classifier.

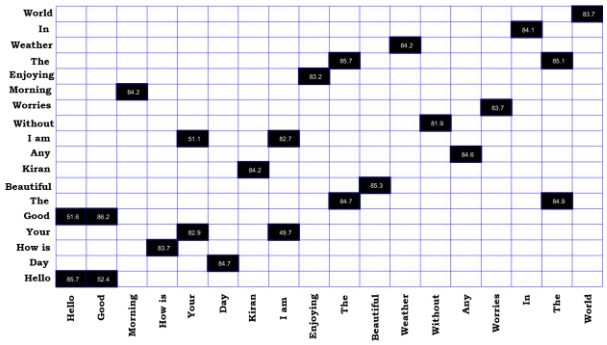

Figure-7. Confusion matrix for Haar features with Adaboost classifier.

Exp-II, uses Local Binary Pattern features with Adaboost classifier. Figure-6(c) show the features projected onto sign frame. The feature vector in this case consisted of unwanted points that does not contribute to the sign. Hence a thresholding mechanism is used to reduce the data size. The confusion matrix generated from LBP features is shown in Figure-8. The average recognition with LBP features is 82.66%.

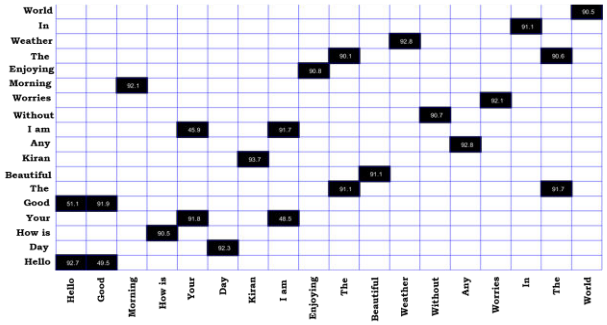

Figure-8. Confusion matrix from LBP features with Adaboost classifier.

Exp-III uses PCA fused Haar - LBP feature represented in Figure-6(d) as the input to the classifier. The confusion matrix is presented in Figure-9. The average recognition from the 151-continuous sign video on 4 different testing samples is around 84.56%. For some individual signs, the recognition is around 95%. Exp - IV uses the novel proposed algorithm where the Haar wavelet features are presented as input to LBP and features are generated as shown in Figure-6(e).

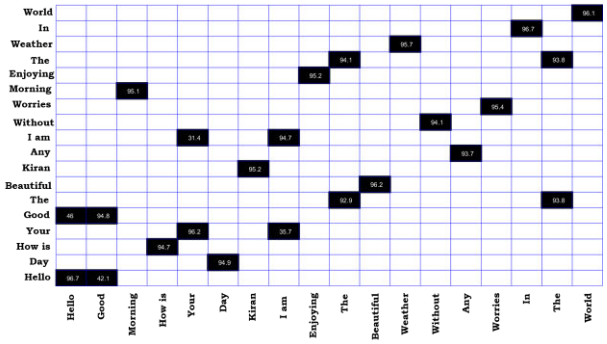

Figure-9. Confusion matrix from PCA fused Haar LBP features with Adaboost classifier.

Figure-10 shows a comparison between single sign features from the proposed DLBP algorithm when there is change in hand positions and shapes along the spatial plane. This shows there is a relative variation in features as the hand position and shape changes in space. Each feature vector is a combination of both spatial information and shape information as mapped onto a Gaussian estimate shown in Figure-10(b).

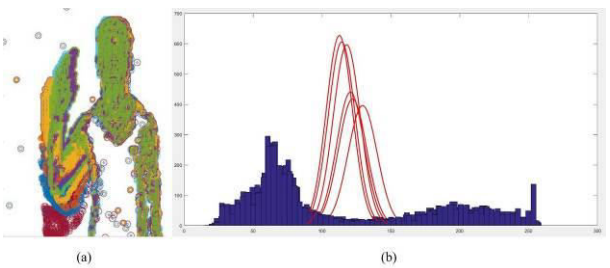

Figure-10. (a) Spatial and location changes in sign and (b) Feature vector mappings on Gaussian estimate.

The confusion matrix from exp-IV using DLBP features is shown in Figure-11. The average recognition in this case is 90.28% for a set of 4 testing vectors. It also shows that DLBP features has reduced the inter dependency on closely matched signs.

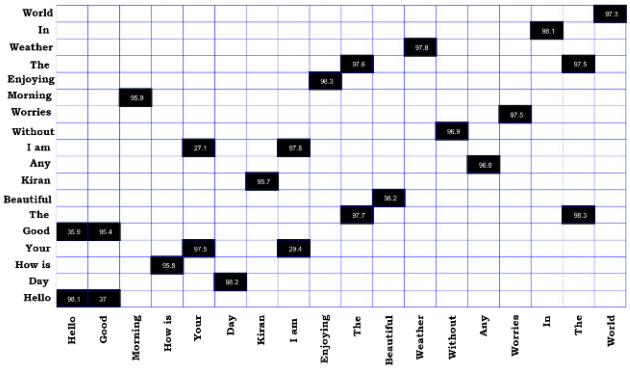

Figure-11. Confusion matrix from DLBP features with Adaboost classifier.

We also tested the classifier with HOG (Histogram of oriented Gradients), SIFT (Scale Invariant Feature Transform) and SURF (Speeded Up Robust Features) with Adaboost classifier. Training vector is made from 50 best features and the same number is used for testing. The average recognition rates were 0.74 for HOG, 0.72 SIFT and 0.7 for SURF. The drop-in classifier performance can be attributed to the poor feature extraction due a large variation in the video frames even though they have same hand pose. HOG, SIFT and SURF features are extracted from the original gray scale video frame. A comparison on the proposed features against the state of the art features on a set of continuous sign videos is shown in figure.12. The comparison shows DLBP features of a sign video is having the ability to classify signs better compared to other modelled features.

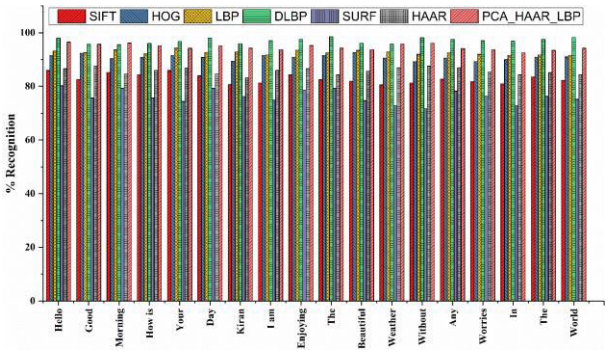

Figure-12. Confusion matrix from DLBP features with Adaboost classifier.

5. CONCLUSIONS

Indian sign language classification is a complex problem in machine vision research. The features representing the signer should focus on the entire human hand shapes and their positions in space. In this work, we proposed a fully automated SLR consisting of signer

identification, extraction, segmentation, feature representation and classification. Saliency based signer identification and extraction helps in reducing the image space. Wavelet reconstructed local binary patterns are used for feature representation preserving local shape content of hands shapes with position vectors. Multi class multi label Adaboost on features of sets of sign video data is the classifier. Multiple experimentations on the video data is tested with 5 different signers. Sign video data is labelled as per the meaning they represent. More action features can be added for representing signer more realistically by elimination backgrounds and blurring artefacts to improve the efficiency of the classifier.

REFERENCES

[1] Parton, Becky Sue. 2006. Sign language recognition and translation: A multidiscipline approach from the field of artificial intelligence. Journal of deaf studies and deaf education. 11(1): 94-101, 2006.

[2] Mitra, Sushmita and Tinku Acharya. 2007. Gesture recognition: A survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 37(3): 311-324.

[3] Raffa Giuseppe, Lama Nachman, and Jinwon Lee. 2017. Efficient gesture processing. U.S. Patent 9,535,506, issued January 3.

[4] Liu, Zhengzhe, Fuyang Huang, Gladys Wai Lan Tang, Felix Yim Binh Sze, Jing Qin, Xiaogang Wang, and Qiang Xu. 2016. Real-time Sign Language Recognition with Guided Deep Convolutional Neural Networks. In Proceedings of the 2016 Symposium on Spatial User Interaction, pp. 187-187. ACM.

[5] Chen, Feng-Sheng, Chih-Ming Fu and Chung-Lin Huang. 2003. Hand gesture recognition using a real-time tracking method and hidden Markov models. Image and vision computing. 21(8): 745-758.

[6] Cavender, Anna, Rahul Vanam, Dane K. Barney, Richard E. Ladner and Eve A. Riskin. 2008. MobileASL: Intelligibility of sign language video over mobile phones. Disability and Rehabilitation: Assistive Technology. 3(1-2): 93-105.

[7] Starner Thad, Joshua Weaver and Alex Pentland. 1998. Real-time American Sign Language recognition using desk and wearable computer based video. IEEE Transactions on Pattern Analysis and Machine Intelligence. 20(12): 1371-1375.

[8] Kushwah, Mukul Singh, Manish Sharma, Kunal Jain, and Anish Chopra. 2017. Sign Language Interpretation Using Pseudo Glove. In Proceeding of International Conference on Intelligent Communication, Control and Devices, pp. 9-18. Springer Singapore.

[9] Kumar, Pradeep, Himaanshu Gauba, Partha Pratim Roy, and Debi Prosad Dogra. 2016. Coupled HMM-based Multi-Sensor Data Fusion for Sign Language Recognition. Pattern Recognition Letters.

[10] Mapari, Rajesh B., and Govind Kharat. 2016. American Static Signs Recognition Using Leap Motion Sensor. In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies, p. 67. ACM.

[11] Auti, Aishwarya, Romalee Amolic, Shubham Bharne, Ankita Raina and D. P. Gaikwad. 2017. Sign-Talk: Hand Gesture Recognition System. International Journal of Computer Applications. 160(9).

[12] Belgacem, Selma, Clément Chatelain, and Thierry Paquet. "Gesture sequence recognition with one shot learned CRF/HMM hybrid model." Image and Vision Computing 61 (2017): 12-21.

[13] Sun, Shiliang, Chen Luo and Junyu Chen. 2017. A review of natural language processing techniques for opinion mining systems. Information Fusion. 36: 10-25.

[14] 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.

[15] Koller Oscar, Jens Forster and Hermann Ney. 2015. Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding. 141: 108-125.

[16] Indian Sign language, empowering the deaf. <http://www.deafsigns.org>.

[17] Gaolin Fang and Wen Gao. 2007. Large Vocabulary Contineous Sign language Recognition Based on Trnsition-Movement Models. IEEE Transaction on Systems, MAN, and Cybernetics. 37(1): 1-9.

[18] A. T.Starner and A. Pentland. 1995. Real-Time American Sign Language Recognition from video

using Hidden Markov Models. Technical Report, MIT Media laboratory Perceptual computing section, Technical Report number. 375.

[19] Ming-Hsuan Yang and Narendra Ahuja. 2002. Extraction of 2D Motion Trajectories and its Application to Hand Gesture Recognition. IEEE Transaction on Pattern Analysis and Machine Intelligence. 24(8): 1061-1074.

[20] M.K. Bhuyan and P.K. Bora. A Frame Work of Hand Gesture Recognition with Applications to Sigb Language. Annual India Conference, IEEE. pp. 1-6.

[21] Rini Akmeliawati, Melanie Po-Leen Ooi and Ye Chow Kuang. 2006. Real-Time Malaysian Sign Language Translation Using Colour Segmentation and Neural Network. IEEE on Instrumentation and Measurement Technology Conference Proceeding, Warsaw, Poland. pp. 1-6.

[22] Nariman Habili, Cheng Chew Lim and Alireza Moini. 2004. 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. 14(8): 1086-1097.

[23] Yu Zhou and Xilin Chen. 2010. Adaptive sign language recognition with Exemplar extraction and MAP/IVFS. IEEE signal processing letters. 17(3): 297-300.

[24] 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.

[25] Sumita, E., Akiba, Y., Doi, T., *et al*. 2003. A Corpus-Centered Approachto Spoken Language Translation. Conf. of the Europ. Chapter of the Ass. For Computational Linguistics (EACL), Budapest, Hungary. pp. 171-174.

[26] Casacuberta F., Vidal E. 2004. Machine translation with inferred stochastic finite-state transducers. Computational Linguistics. 30(2): 205-225.

[27] Och J., Ney H. 2003. A systematic comparison of various alignment models. Computational Linguistics. 29(1): 19-51.

[28] Koehn P. 2004. Pharaoh: a beam search decoder for phrase-based statistical machine translation models. AMTA.

[29] P.V.V. Kishore, P. Rajesh Kumar, E. Kiran Kumar, S.R.C. Kishore. 2011. Video Audio Interface for Recognizing Gestures of Indian Sign Language. International Journal of Image Processing (IJIP), CSC Journals, Kualalumpur, Malaysia. 5(4): ISSN: 1985-2304, pp. 479-503.

[30] Kishore P. V. V., *et al*. 2015. 4-Camera model for sign language recognition using elliptical Fourier descriptors and ANN. Signal Processing And Communication Engineering Systems (SPACES), 2015 International Conference on. IEEE.

[31] Kishore P. V. V., A. S. C. S. Sastry and A. Kartheek. 2014. Visual-verbal machine interpreter for sign language recognition under versatile video backgrounds. Networks & Soft Computing (ICNSC), 2014 First International Conference on. IEEE.

[32] 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.

[33] Anil Kumar, D., Kishore, P. V. V., Prasad, M.V. D., Raghava Prasad, Ch. 2016. Fuzzy Classifier for Continuous Sign Language Recognition from Tracking and Shape Features. Indian Journal of Science and Technology. Vol. 9.

[34] P. V. V. Kishore, M. V. D. Prasad, C. R. Prasad and R. Rahul. 2015. 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN. 2015 International Conference in Signal Processing and Communication Engineering Systems (SPACES), doi: 10.1109/SPACES.2015.7058288, pp. 34-38.

[35] Kishore PVV, Rajesh Kumar P. 2012. A video based Indian Sign Language Recognition System (INSLR) using wavelet transform and fuzzy logic. International Journal of Engineering and Technology. 4(5): 537-42.

[36] Inthiyaz, Syed, B. T. P. Madhav and P. V. V. Kishore. 2017. Flower segmentation with level sets evolution controlled by colour, texture and shape features." Cogent Engineering. 4(1): 1323572.

[37] Rätsch Gunnar, Takashi Onoda and K-R. Müller. 2001. Soft margins for AdaBoost. Machine learning. 42(3): 287-320.

[38] Wu, Bo, Haizhou Ai, Chang Huang and Shihong Lao. 2004. Fast rotation invariant multi-view face detection based on real adaboost. In Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, pp. 79-84. IEEE.

[39] Zhu Ji, Hui Zou, Saharon Rosset and Trevor Hastie. 2009. Multi-class adaboost. Statistics and its Interface. 2(3): 349-360.

[40] Qi, Chengming, Zhangbing Zhou, Yunchuan Sun, Houbing Song, Lishuan Hu and Qun Wang. 2017. Feature selection and multiple kernel boosting framework based on pso with mutation mechanism for hyperspectral classification. Neuro computing. 220: 181-190.

[41] Patel Chirag I., Sanjay Garg, Tanish Zaveri, Asim Banerjee and Ripal Patel. 2016. Human action recognition using fusion of features for unconstrained video sequences. Computers & Electrical Engineering.

[42] Wang Jin, Mary She, Saeid Nahavandi and Abbas Kouzani. 2010. A review of vision-based gait recognition methods for human identification. In Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on, pp. 320-327. IEEE.

[43] Ben-Che, Mirela and Craig Gotsman. 2008. Characterizing Shape Using Conformal Factors. In 3DOR. pp. 1-8.

[44] Wu Bo, Haizhou Ai, Chang Huang and Shihong Lao. 2004. Fast rotation invariant multi-view face detection based on real adaboost. In Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, pp. 79-84. IEEE.

[45] Zhu Ji, Hui Zou, Saharon Rosset and Trevor Hastie. 2009. Multi-class adaboost. Statistics and its Interface. 2(3): 349-360.

[46] Qi, Chengming, Zhangbing Zhou, Yunchuan Sun, Houbing Song, Lishuan Hu and Qun Wang. 2017. Feature selection and multiple kernel boosting framework based on pso with mutation mechanism for hyperspectral classification. Neurocomputing. 220: 181-190.

[47] Khare, Manish, Rajneesh Kumar Srivastava and Ashish Khare. 2017. Object tracking using combination of daubechies complex wavelet transform and Zernike moment. Multimedia Tools and Applications. 76(1): 1247-1290.