Segmentation, Tracking And Feature Extraction For Indian Sign Language Recognition

Divya S , Kiruthika ,S Nivin Anton A L and Padmavathi S

Student, Department of Computer Science & Engineering, Amrita University

ABSTRACT

Sign Language is a means of communication between audibly challenged people. To provide an interface between the audibly challenged community and the rest of the world we need Sign Language translators. A sign language recognition system computerizes the work of a sign language translator. Every Sign Language Recognition (SLR) System is trained to recognize specific sets of signs and they correspondingly output the sign in the required format. These SLR systems are built with powerful image processing techniques. The sign language recognition systems are capable of recognizing a specific set of signing gestures and output the corresponding text/audio. Most of these systems involve the techniques of detection, segmentation, tracking, gesture recognition and classification. This paper proposes a design for a SLR System.

KEYWORDS

Indian Sign Language (ISL), Sign Language Recognition Systems (SLR Systems), Skin Based Segmentation, RGB, HSV, YCbCr, Kalman Filter, SIFT.

1. INTRODUCTION

Sign Language Translators act an interface of communication between the audibly challenged and the rest of the world. Providing an interface between the communities in Indian Sign Language (ISL) is very challenging issue because of the following reasons

(1) The ISL Signs have not been standardized as of date.
(2) The ISL Signs vary from region to region and hence we have various dialects of ISL.
(3) Every Signer does the same gesture in a different ways.
(4) Information regarding the signs and gestures are multichannel.

The Central Institute ISL Society is currently working in collaboration with Ramakrishna Mission to standardize the ISL Signs.

ISL words are categorized into 4 types as- (1) Feelings (2) Descriptions (3) Actions and (4) Non-Manual Actions. ISL grammar omits the usage of articles such as (a, an, the) and also does not include tense forms. The sentence structure of an ISL sentence (SOV) – Subject Object Verb- very much different from the sentence structure of English Language (SVO) – Subject Verb Object and hence ISL and English Language aren't verbatim convertible. The term sign and gesture are used interchangeably throughout the paper. The Signs are represented by the HamNoSys* System of Notation for Sign Language. Most of the signs in ISL involve dynamic movement of both the hands and non-manual gestures. SLR systems are simple yet trained intelligent systems used for converting sign language into text by recognizing the head and hand gestures. SLR systems have been classified into two types based on the approach- (1) Data Glove

Based Method and (2) Vision Based Method. Data Glove Based Approach simplifies the recognition of the gestures but involves complicated hardware. The Vision Based Approach is a very user friendly and does not involve complicated hardware usage and hence its most widely used method in SLR Systems

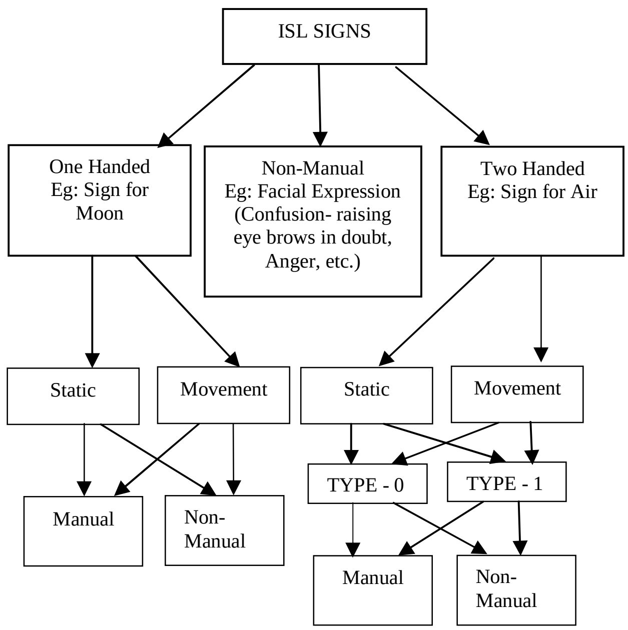

Figure: 1 – Taxonomy of ISL Signs [1]

The paper is organized as follows; Section II provides the Survey of the various SLR Systems. Section III gives a detail description of the proposed SLR System. The Experimental Results are described in the Section IV. Finally, Section V provides a few concluding remarks and lists out the acknowledgments.

2. LITERATURE SURVEY

Most of the SLR systems today use the Vision Based – Skin Colour Approach since the gestures to be identified depend gesturing of the hands and face. The skin based filtering help to extract the necessary components from the other coloured pixels and also helped to extract objects from their background. [2-10]. Most of the skin based methods use the pixel by pixel method and these methods help to determine whether the given pixel intensity is a skin pixel or not. Prior to these methods, statistical models like Gaussian Mixture Model [8], Single Gaussian Model [9] and Histogram Based Approaches [10] were used. Many researchers also made use of Coloured Gloves to simplify the process of segmentation [12].

Paulraj et al [2] built an SLR System on a Skin colour based approach. After segmentation the frames were converted into binary images. During feature extraction, the areas of the three segmented areas were calculated and noted down as discrete events. Since each gesture involves different segmented areas, the DCT coefficients of the three segmented regions were taken as the features. These features were given as input to the Neural Network (NN) Classifier. The NN does the gesture classification and plays the corresponding audio signal. The system has a minimum and maximum classification rate of 88% and 95% respectively.

Joyeeta Singha and Karen Das [3] developed a system to recognize alphabets. The regions of interest were cropped and the Eigen values and vectors of these regions were used as features and also as input to the classification system. A novel approach called the Eigen Value Weighted Euclidean Distance was used to classify the gestures. The system has a success rate of 97%.

An SLR System was proposed by Kishore Kumar et al [4] to automatically recognize gestures of words and convert them to text or audio format. A combination of wavelet transforms and image fusion techniques was used during the process of segmentation. Edge Based tracking methods was used to track the gesture movements. Elliptical Fourier descriptors were taken as features and it is given as input to the fuzzy system, which was used to classify and recognize the gestures. The system has a success rate of 91%.

Deepika Tewari and Sanjay Kumar Srivastava proposed a method for the recognition of Indian Sign Language [5] in which gesture frames were taken with a digital camera. Features were extracted using 2D Discrete Cosine Transform (2D-DCT) for each detected region, and feature vectors were formed from the DCT coefficients. Self-organizing map (SOM) using an unsupervised learning technique in Artificial Neural Network (ANN) was used to classify DCT-based feature vectors into groups and to decide if the gesture performed is either present or absent in the ISL database. The recognition rate achieved was 80%.

Joe Naoum-Sawaya et al proposed a system for American Sign Language Recognition [6]. The hands were detected by performing motion detection over a static background. After background subtraction, the image is segmented using skin colour based thresholding methods. Morphological operations were done to improve the robustness of the segmentation results. The CAMSHIFT (Continuously Adaptive Mean- SHIFT) algorithm is used to draw a box about the region of the hand to simplify template matching. The gesture is recognized by the template with the highest matching metric. Each of the dynamic gestures is coded as a series of directions and static gestures. The accuracy for this system is 96% in daylight with distinct backgrounds.

Jaspreet Kaur et al uses modified SIFT algorithm for American Sign Language[7].A modified method of comparability measurement is introduced to improve the efficiency of the SIFT algorithm. This system is highly robust to scale difference, rotation by any angle and reflection and is 98.7% accurate for recognizing gestures. Typically every SLR system involves various techniques in the following categories

i. Segmentation or Detection
ii. Tracking the Gestures
iii. Feature Extraction and Gesture Recognition
iv. Gesture Classification

We have provided a detailed survey of the methods that can used for the above categories in [28][29][30].The presence or absence of any of the above mentioned categories depends on the input dataset. The successive section provides a detailed account of our proposed system designed to recognize ISL signs.

3. PROPOSED SLR SYSTEM

The section elucidates the detailed description of our proposed system to recognize Indian Sign Language words. The dataset chosen involves mostly two handed movement gestures without any non-manual markers. The only regions of interest are the two hands and the face region. The system requires a constant background and also uniform lighting conditions. Like every other SLR system, our system also involves the stages of Detection or Segmentation, Morphological operations to remove noise, Tracking the motion of the regions, Feature Extraction, Gesture Recognition, Gesture Classification and System Learning. The input video is taken from a digital camera and converted into frames. The technique of segmentation here uses the skin colour based method and hence extracts the skin colour regions.

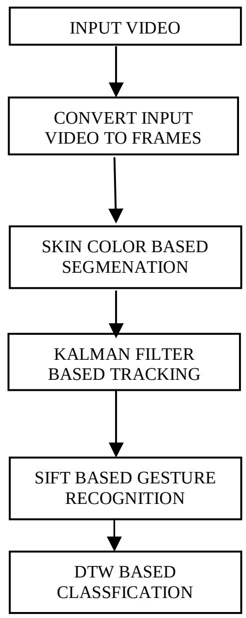

Figure 2: Flow Diagram for the Proposed System

The system requires the signer to wear a dark long sleeved shirt except shades of skin colour. The skin colour based segmentation is a fairly simple method with the only decision to be made on which colour space to use. The result of this method depends dominantly on the chosen colour space. The RGB colour space is an additive colour space and also used one of the most commonly used colour space. The RGB space has high correlation between channels, lot of non-uniformity and also there is a dominant mixing of the chromaticity and the luminance information. [14] Hence the use of the RGB space alone is not suitable for colour based recognition [15]. To overcome this problem the normalized RGB has been introduced to obtain the chromaticity information for more robust results. [16][17][18][19][28].

$$r^N = \frac{R}{R+G+B} \qquad \text{Eq. (1)}$$

$$g^N - \frac{G}{R+G+B} \qquad \text{Eq. (2)}$$

$$b^N = \frac{B}{R+G+B} \qquad \text{Eq. (3)}$$

In the above equations, $r^N, g^N \; and \; g^N$ represent the normalized values of R, G and B respectively. These normalized values will satisfy the Eq. (4).

$$r^N + g^N + b^N = 1 \qquad \text{Eq. (4)}$$

The following are the equations used for the conversion of pixel intensities from RGB to HSV.

$$Max = Maximum\ (r^N, g^N, b^N) \qquad \text{Eq. (5)}$$

$$Min = Minimum(r^N, g^N, b^N) \qquad \text{Eq. (6)}$$

$$H = \begin{cases} 0 & \text{If Max} = \text{Min} \\ 60 * \frac{g^N - b^N}{Max - Min} & \text{If Max} = r^N \\ 60 * \frac{b^N - r^N}{Max - Min} & \text{If Max} = g^N \\ 60 * \frac{r^N - g^N}{Max - Min} & \text{If Max} = b^N \end{cases} \qquad \text{Eq. (7)}$$

$$S = \begin{cases} 0 & \text{If Max} = 0 \\ \frac{Max - Min}{Max} & \text{Otherwise} \end{cases} \qquad \text{Eq. (8)}$$

$$V = Max \qquad \text{Eq. (9)}$$

The following matrix is used to convert intensities from RGB colour space to YCbCr Colour Space.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad \text{Eq. (10)}$$

In this chosen method [20] for skin colour segmentation the pixel intensity if tested for skin colour in HSV, RGB and YCbCr Colour space. Only if the pixel is found to be a skin colour pixel in all the colour spaces it is taken as a skin colour pixel or else it is identified as background pixel. The thresholding conditions for the skin colour classification in the various colour spaces are as follows:

In the RGB space, the following rule put forward by Peer et al [21], determines the skin colour under uniform daylight.

$$\left.\begin{array}{l} (R > 95)\ \&\&\ (G > 40)\ \&\&\ (B > 20) \quad \text{AND} \\ \quad \{MAX(R,G,B) - MIN(R,G,B)\} > 15 \quad \text{AND} \\ (|R - G| > 15) AND\ (R > G) AND\ (R > B) \end{array}\right\} \qquad \text{Eq. (11)}$$

For skin colour under flashlight, a lateral rule for detection is given by

$$\left.\begin{array}{l} (R > 220)\ AND\ (G > 210)\ AND\ (B > 170) \quad \text{AND} \\ (|R - G| \leq 15)\ AND\ (R > B)\ AND\ (G > B) \end{array}\right\} \qquad \text{Eq.(12)}$$

To take note of both the conditions, the above two rules are combined using the logical operator, OR.

In YCbCr space, the following thresholding conditions are applied.

$$\left.\begin{aligned} &Cr \leq 1.5862 * Cb = 20 \\ &Cr \geq 0.3448 * Cb + 76.2069 \\ &Cr \geq -4.5652 * Cb + 234.5652 \\ &Cr \leq -1.15 * Cb + 301.75 \\ &Cr \leq -2.287 * Cb + 432.85 \end{aligned}\right\} \qquad \text{Eq. (13)}$$

In [21] the thresholds for the YCbCr was improved by including the parameters for Y and Cb.

$$Y > 80 \; AND \; 85 < Cb < 135 \;\; AND \;\; 135 < Cr < 180 \qquad \text{Eq. (14)}$$

In the HSV space, the conditions for skin colour are:

$$\left.\begin{aligned} &H < 25 \\ &H > 230 \\ &S > 38 \\ &S < 250 \\ &V > 51 \\ &V < 242 \end{aligned}\right\} \qquad \text{Eq.(15)}$$

The above two condition are combined using a Logical operator OR. If the pixel intensity falls under all the above threshold conditions, then the pixel intensity will be classified as a skin colour intensity.

The results obtained from the segmentation are further acted upon by morphological operations are used to eliminate the presence of noise in the segmentation results. The operations like erode, dilate, open and close are done to obtain needed results.

The contours of the segmented regions are obtained and the smaller contours are eliminated by the sorting the areas of the contours and taking the contours whose areas are significantly the largest three. Taking into the consideration that the largest contour will the head, we fix the position of the head. Based on the position of the head, the other two contours can be identified as left and right hand. After the process of segmentation, bounding boxes are drawn around the three segmented regions, obtained by the above mentioned process.

The flow diagram for the segmentation process is as follows:

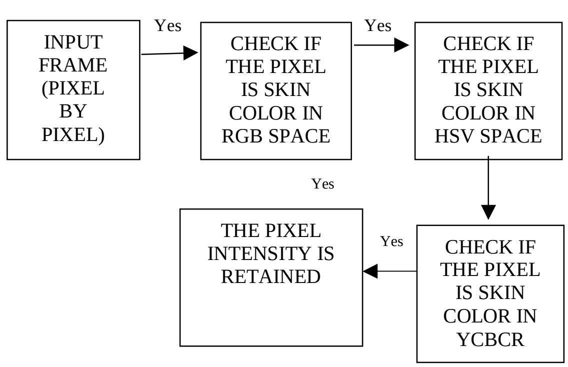

Figure 3: Steps in Segmentation

In cases of occlusion of hands or head & hand, there might be only one or two contours visible depending on the sign and when occlusion occurs, the Kalman Filter begins to predict the position of the regions.

For each of the regions, the following sequence is done.

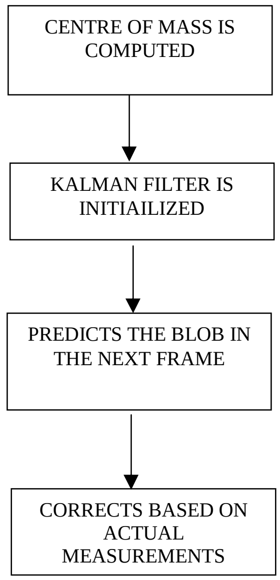

Figure 4: Steps in Tracking

The parameters of the bounding box are used to give input to the Kalman Filter [22] for initiating the tracking process.

The principle of the Kalman Filter [23] involves the prediction of the location of the head and hand regions in the consecutive frames based on the current state and previous state estimates. Kalman Filter takes as input - the measurements of the present state and does a prediction of the consecutive states.

Kalman Filter is optimal estimator of the successive states which has two main methods predict and correct methods. The predict method does an estimate of the next state and the correct method optimizes the system taking into consideration the actual measurements of the regions. The prediction step uses the state model to predict the next state of the system [13]

$$\bar{X}^t = DX^{t-1} + W \qquad \text{Eq. (16)}$$

$$\bar{\sum}^t = D\sum^{t-1}D^T + Q^t \qquad \text{Eq. (17)}$$

$\bar{X}^t$ and $\bar{\sum}^t$ are the state and covariance prediction at time t. D represents the state transition matrix which define the relationship between the state variable at time t and t-1. The matrix Q is the covariance of the white noise W.

The correction step uses the actual observations of the object's current position Z^t to update the state of the object.

$$K^t = \bar{\sum}^t M^t [M \bar{\sum}^t M^t + R^t \ ]^{-1} \qquad \text{Eq. (18)}$$

$$X^t = \bar{X}^t + K^t[Z^t - M\bar{X}^t] \quad \text{Eq. (19)}$$

$$\Sigma^t = \bar{\Sigma}^t - K^t M \bar{\Sigma}^t \quad \text{Eq. (20)}$$

K is the Kalman Gain and Eq. (17) is the Ricatti equation used for the propagation of the state models. X^t is the update state. In Eq. (18), the term $[Z^t - M\bar{X}^t]$ is called the innovation.

The three regions are capable of movement in the 2D space and hence each region at the (x, y) position with velocities (v_x, v_y) follow the equations:

$$x^p = x + v_x t + \frac{1}{2}at^2 \quad \text{Eq. (21)}$$

$$y^p = y + v_y t + \frac{1}{2}at^2 \quad \text{Eq. (22)}$$

$$v_x = v_{x|t-1} + a(t-1) \quad \text{Eq. (23)}$$

$$v_y = v_{y|t-1} + a(t-1) \quad \text{Eq. (24)}$$

x^p and y^p represent the change in the positions in the X and Y directions respectively. The terms $v_{x|t-1}$ and $v_{y|t-1}$ represent the 'x' and 'y' velocities at time (t-1). The regions are represented by the state vector which holds information about the x position, y position, horizontal and vertical velocities. v_x & v_y represent the velocities in the x and y direction at time't'.

The state vector $\hat{x}$ is represented as follows

$$\hat{x} = (x^p, y^p, v_x, v_y) \quad \text{Eq. (25)}$$

Since we use a only a four dimensional vector for representing the state, the transition matrix can be simply given by

$$D = 0 \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{Eq. (26)}$$

The three regions can be easily tracked by computing the Centre of Mass.[29] The Centre of Mass of the three regions are the initial parameters given to the three objects of Kalman Filter. The initial detections of the regions are used to initialize the Kalman Filters for the three regions. The next step involves predicting the new locations of the three regions based on the previous state parameters. The predict and correct methods of the Kalman Filter are used in sequence help reduce the noise in the system. The predict method by itself is used to estimate the location of a region when it is occluded by another region. The correct method which corrects and optimizes the prediction mechanism is based on the actual measurements. The steps of the Filter are recursive till required. This method provides the trajectory of motion of the hands and head. The Kalman Filter Based Multiple Object Tracking is robust to occlusions and movements involving rotation and scaling. The tracking step is very essential when the occlusion of the two or three regions occurs, since the tracker continues to track the regions even if region is not in the field of view.

The next step in the process of sign language recognition is the Gesture Recognition. For this process, we are using Scale Invariant Feature Transform (SIFT), to find the gestures and its invariant to translation, rotation, scaling and its partially invariant to illumination. The image feature recognition is performed through four phases [6] of which the various phases are (1) Scale space local extrema detection (2) Key Point Localization (3) Orientation Assignment (4) Key Point Descriptor. In the first phase the key points from various orientations are obtained. The Scale Space Function is available for this phase. [24]

$$D(x,y,\sigma) = G(x,y,k\sigma) \quad G(x,y,\sigma) * I(x,y) \qquad \text{Eq. (27)}$$

where $I(x,y)$ is the Input Image, $G(x,y,\sigma)$ is the variable Gaussian Scale. Scale Space Extrema $D(x,y,\sigma)$ is found by difference between two images, one with k times the other.

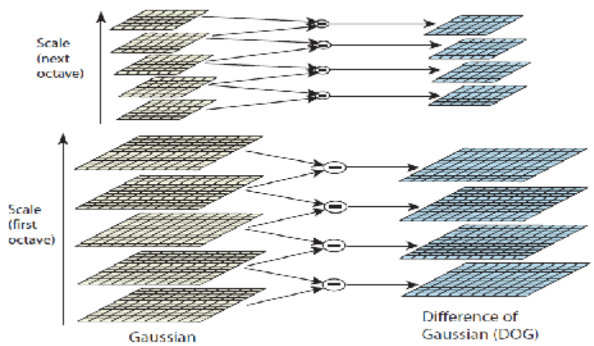

Figure 5: Difference of Gaussian (DoG)[25]

To obtain the local minima and maxima we compare each key point with eight neighbours on the same scale and nine neighbours on scales above and below it. The next phase involves fitting of the key points in the three degree quadratic functions. This function is obtained from the second order Taylor Expansion. The local extrema with lower contrasts will not take into consideration because they are sensitive to noise. The key points below the threshold level are also not taken in account by the system.

The next phase involves the orientation assignment where the main orientation is assigned to each feature based on local image gradient. For each pixel around the key point the gradient magnitude and the orientation are computed. The following equations are used to find the magnitude and orientation [25]

$$M(x,y) = \sqrt{\left(L(x+1,y) - L(x-1,y)\right)^2 + \left(L(x,y+1) - L(x,y-1)\right)^2} \qquad \text{Eq. (28)}$$

The above equation is used to calculate the magnitude $M(x,y)$ of the detected key points.

$$\Theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \qquad \text{Eq. (29)}$$

In the above equation, $\Theta(x, y)$ gives the orientation of the key point. Eq (28) is to calculate the orientation of the detected key points. The orientation and magnitude of the key points are stored and used in the further process of Gesture Classification.

The next phase of gesture recognition is the key point descriptor phase. In this the local image gradients are measured at a selected scale around each key point.
The following diagram [30] specifies the steps in the gesture recognition process.

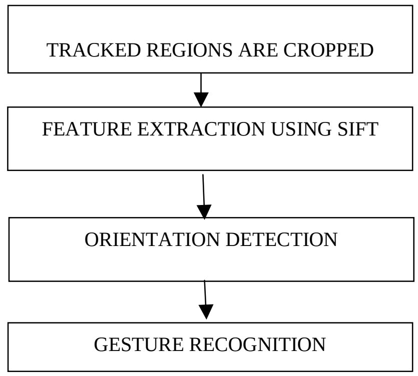

Figure 6: Steps in Gesture Recognition

The final phase of the Sign Language Recognition System is the classifier system. The Classification System uses the concept for classifying the gestures. A number of Classification Systems have been used extensively used.

4. EXPERIMENTAL RESULTS

The designed system has been tested on 25 different signs of ISL. The domain words are restricted to child learning. The signs have been performed under varying conditions with 15 different signers. Each of these test cases have been performed at varying lighting and under complex background. A sample of our dataset is shown in Fig - 7.

Figure 8: Gestures for the word (Girl)

The Segmentation algorithm is applied on the input signs are the results obtained the Fig.9 (a) & (b) are the skin colour representations of the segmentation results. The binary skin color segmentation is shown in the Fig- 10 (a) & (b).

(a) (b)

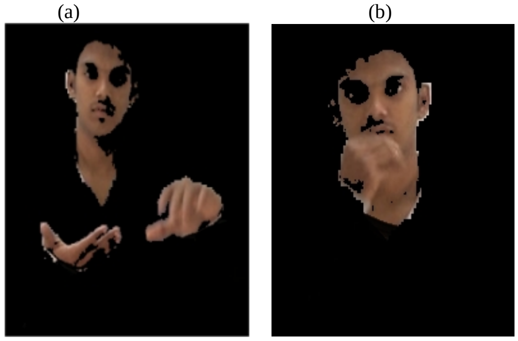

Figure 9: Skin Colour Regions – Segmentation Results

(a) (b)

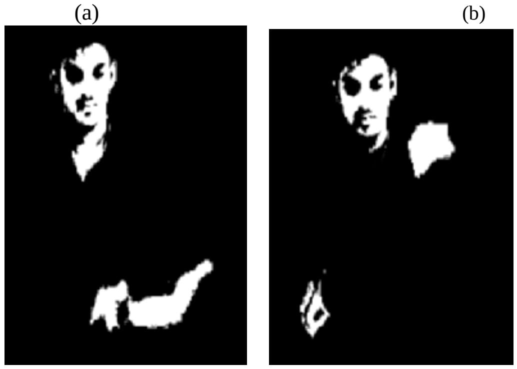

Figure 10: Skin Colour Segmentation – Binary Threshold Results

Skin colour segmentation fails to produce proper segmentation results when the signer wears dresses of shades similar to skin colour. The results obtained in the presence of other skin colour objects in the scenario and also when the signer is clothed in shades of skin colour are shown in Fig – 11(a) & (b). Hence it can be inferred that system produces erroneous results in the presence of other skin coloured objects.

(a) (b)

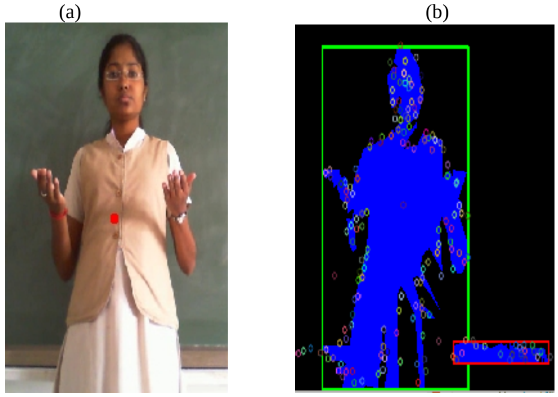

Figure 11: (a) Input Frame (b) Erroneous Results

The results of segmentation, tracking and gesture recognition also varies in the presence of variable lighting conditions. When the signing of the word is done under variable lighting conditions, the algorithm fails to find the necessary results as show in the Fig – 12(a) & (b).

(a) (b)

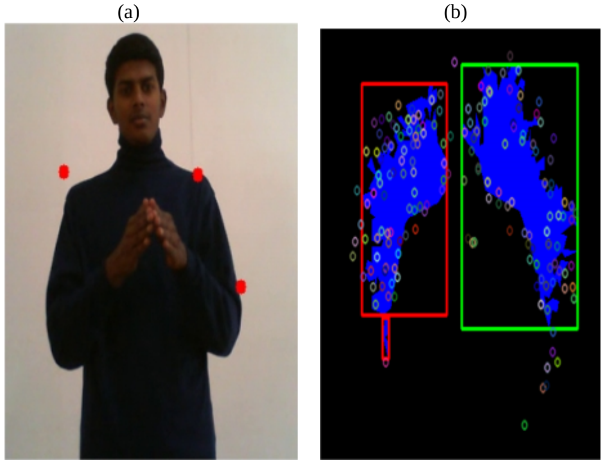

Figure 12: Results – Non-Uniform Lighting Conditions (a) Input Frame (b) Output

The system works perfectly well under uniform lighting condition. The system constraints also require the signer to wear long sleeved dark clothing for the algorithm to produce the required results. The results are as shown in figure in Fig-13(a) & (b). The Fig-14(a) & (b) represent the input frame and the output where the system handles the occlusion of the hands.

(a) (b)

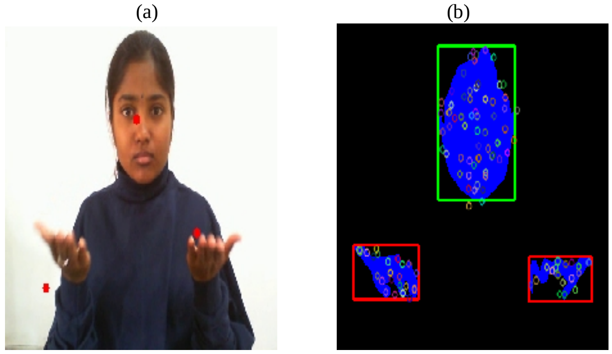

Figure 13: Results – Uniform Lighting (a) Input Frame (b) Output

(a) (b)

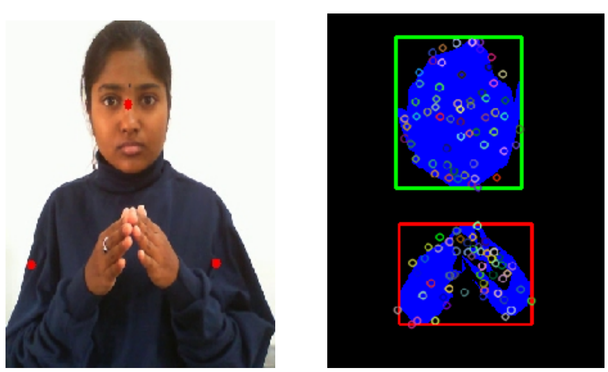

Figure 14: Results – Occlusion under Uniform Lighting (a) Input Frame (b) Output

(a) (b)

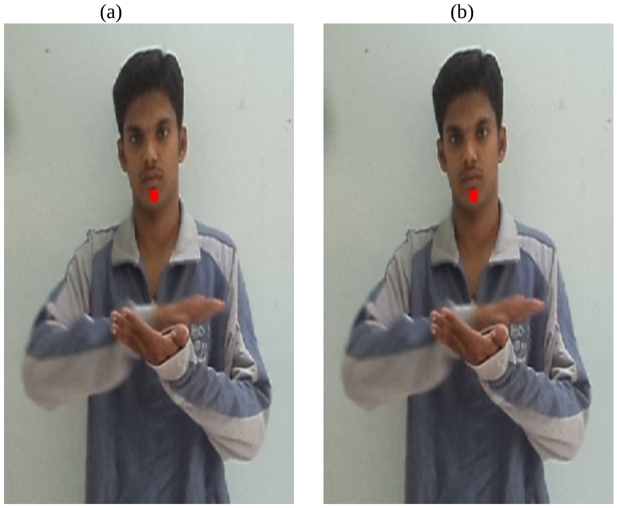

Figure 15: Results (a) Input Frame (b) Output

Real time sign language recognition under varying conditions of lighting causes errors in the segmentation of the regions and hence it is inferred that the further processes of tracking, gesture recognition and classification depend on the results of segmentation. The results of segmentation are refined using morphological operations.

The segmentation results for the three colour spaces are shown below separately. Figure-16 represents the HSV component separately. Figure-17 shows the output for RGB component. Figure-17 represents the YCbCr.

(a) (b)

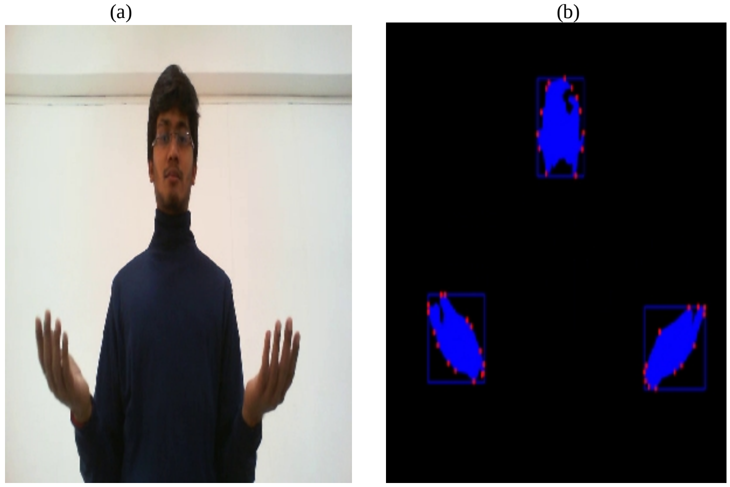

Figure -16 (a) Input (b) HSV component

(a) (b)

Figure: 17 (a) Input (b) RGB Component

(a) (b)

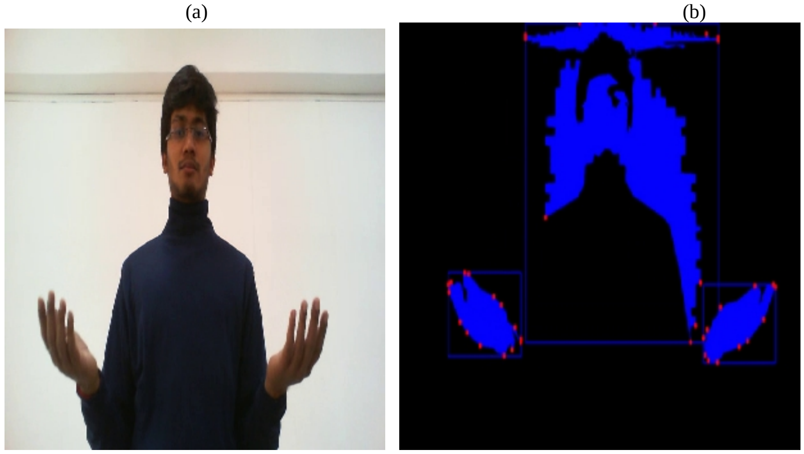

Figure: 18 (a) Input (b) YCbCr Component

Figure-19 represents the result of tracking using Particle Filter and Keypoints plotting using SIFT Algorithm. The red dots represents the tracking points of Hands and head separately. The other colour dots shows the key points.

(a) (b)

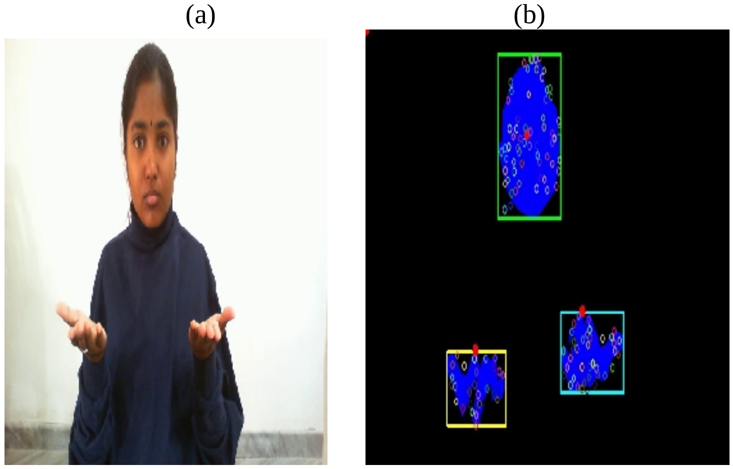

Figure: 19(a) Input (b) Tracking and SIFT

5. CONCLUSIONS

Our SLR requires the signer to perform the signing under constant illumination and also requires the signer to wear a long sleeved attire which can be constraint to a signer to do the same under natural constraints. The system developed uses a simple parametric pixel based segmentation method which can be further improved using system training methods. The performance of tracking has a lot of scope of improvement. The Gesture Recognition mechanism can be replaced by Advanced SIFT for improved accuracy. The input dataset can be further extended and the classifying system can be further trained by providing more positive and negative samples.

ACKNOWLEDGEMENTS

We take this opportunity to express our sincere thanks to “Ramakrishna Mission Vivekananda University, Coimbatore Campus FDMSE- Indian Sign Language” for providing us valuable information regarding the Signs of Indian Sign Language. We also express our deepest gratitude to Ms. Poongothai, ISL Interpreter for helping us with our dataset.

REFERENCES

[1] Tirthankar Dasgupta, Sambit Shukla, Sandeep Kumar, Synny Diwakar, Anupam Basu.”A Multilingual Multimedia Indian Sign Language Dictionary Tool”, the 6th Workshop on Asian Language Resources, 2008.

[2] Paulraj M P, Sazali Yaacob, Hazry Desa, Hema C.R., “Extraction of Head & Hand Gesture Feature for Recognition of sign language”, International Conference on Electronic Design, Penang, Malaysia, December 1-3, 2008.

[3] Joyeeta Singha & Karen Das, “Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique”, International Journal of Advanced Computer Science and Applications, Vol. 4, No. 2, 2013.

[4] P.V.V Kishore, P.Rajesh Kumar , E. Kiran Kumar & S.R.C Kishore , “Video Audio Interface for Recognizing Gestures of Indian Sign Language”, International Journal of Image Processing , Volume 5, Issue 4, 2011, pp. 479-503.

[5] Deepika Tewari, Sanjay Kumar Srivastava, “A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self – Organizing Map Algorithm”, International Journal of Engineering & Advanced Technology (IJEAT), Vol. 2, Issue-2, December 2012.

[6] Joe Naoum-Sawaya, Mazen Slim, Sami Khawam and Mohamad Adnan Al-Alaoui, ”Dynamic System Design for American Sign Language Recognition”, Electrical and Computer Engineering Department, American University of Beirut, Beirut, Lebanon.

[7] Jaspreet Kaur, Navjot Kaur, “Modified SIFT Algorithm for Appearance Based Recognition of American Sign Language”, IJCSET, Vol. 2, Issue 5, May 2012.

[8] J.-C. Terrillon, M. David, and S. Akamatsu,,”Automatic Detection of Human Faces in Natural Scene Images by use of a Skin Color Model and of Invariant Moments”,Proc.Int.Conf.AFGR’98, Nara, Japan, pp. 112-117,1998.

[9] S.J. McKenna, S. Gong, and Y. Raja, “Modeling Facial Color and Identity with Gaussian Mixtures”, Pattern Recognition, 31(12), pp. 1883-1892, 1998.

[10] R. Kjeldsen and J. Kender, “Finding Skin in Color Images”, Proc.Int. Conf. AFGR’96, Killington, Vermont, pp. 312-317, 1996.

[11] A. A. Argyros, M. I. A. Lourakis: "Real-Time Tracking of Multiple Skin-Colored Objects with a Possibly Moving Camera", ECCV, 2004.

[12] Rini Akmeliawati , Melanie Po-Leen Ooi and Ye Chow Kuang, “Real Time Malaysian Sign Language Translation using Color Segmentation and Neural Network”,IMTC 2007- Instrumentation and Measurement Technology Conference Warsaw , Poland ,1-3, May 2007.

[13] Yilmaz, A., Javed, O., and Shah, M. 2006. “Object tracking: A survey”. ACM Computing. Survey. 38, 4, Article 13 (Dec. 2006), 45 pages. DOI = 10.1145/1177352.1177355 http://doi.acm.org/10.1145/1177352.1177355

[14] V. Vezhnevets, V. Sazonov, A. Andreeva," A survey on pixel-based skin color detection techniques,"GRAPHICON03, 2003, pp. 85–92.

[15] Manresa, C., Varona, J., Mas, R., Perales, F.J.,"Real-time hand tracking and gesture recognition for human-computer interaction", ELCVIA (5), No. 3, 2005, Pages: 96-104.

[16] Beetz, M., B. Radig, and M. Wimmer. "A person and context specific approach for skin color classification", 18th International Conference on Pattern Recognition, 2006 (ICPR 2006). 2006. Hong Kong.

[17] Soriano, M., et al., "Skin detection in video under changing illumination conditions", 15th International Conference on Pattern Recognition 2000. 2000. Barcelona.

[18] Kawato, S. and J. Ohya., "Automatic skin-color distribution extraction for face detection and tracking", 5th International Conference on Signal Processing Proceedings 2000 (WCCC-ICSP 2000). 2000. Beijing.

[19] Park, J., et al., "Detection of human faces using skin color and eyes", 2000 IEEE International Conference on Multimedia and Expo 2000 (ICME 2000). 2000. New York, NY.

[20] Nusirwan Anwar bin Abdul Rahman, Kit Chong Wei and John, "RGB-H-CbCr Skin Color Model for Human Face Detection", MMU International Symposium on Information & Communications Technologies (M2USIC), 2006.

[21] G.Kukharev, A.Novosielski, "Visitor identification – elaborating real time face recognition system," Proceedings of the 12th Winter School on Computer Graphics (WSCG) , Plzen, Czech Republic , pp. 157-164, February , 2004.

[21] P. Peer, J. Kovac, F. Solina, "Human Skin Color Clustering for Face Detection", EUROCON1993, Ljubljana, Slovenia, pp. 144-148, September 2003.

[22] R.E Kalman, "A new approach to linear filtering and prediction problems", Transactions of the ASME, Ser. D., Journal of Basic Engineering, 82, 34-45. (1960).

[23] Welch, G., Bishop, G.: "An Introduction to the Kalman Filter", ACM SIGGRAPH 2001, Course 8, available at http://www.cs.unc.edu/∼welch/Kalman/ (2001).

[24] Pallavi Gurjal , Kiran Kunnur, "Real Time Hand Gesture Recognition Using SIFT", International Journal of Electronics and Electrical Engineering, Volume 2 Issue 3 ,March 2012.

[25] Sakshi Goyal, Ishita Sharma, Shanu Sharma, "Sign Language Recognition System for Deaf and Dumb People, "International Journal of Engineering Research &Technology, Vol. 2 Issue 4, April – 2013.

[26] Sakoe. H and Chiba, S. "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Trans. on Acoust Speech, and Signal Process. ASSP 26, 43-49 (1978).

[27] Bautista, M., Hernández-Vela, A., Ponce, V., Perez-Sala, X., Baro´, X., Pujol, O., Angulo, C., Escalera, S., "Probability-based dynamic time warping for gesture recognition on RGB-D data". In: International Workshop on Depth Image Analysis. Tsukuba Science City, Japan (2012)

[28] Padmavathi, S., Nivin Anton, AL., "Survey of Skin Color Segmentation Techniques", International Journal of Engineering Research & Technology, Vol. 3, Issue 2, February- 2014.

[29] Padmavathi, S., Divya, S., "Survey on Tracking Algorithms", International Journal of Engineering Research & Technology, Vol. 3, Issue 2, February- 2014.

[30] Padmavathi, S., Kiruthika, S., Navin Kumar, PI. , "Survey on Hand Gesture Recognition", International Journal of Engineering Research & Technology, Vol. 3, Issue 2, February- 2014.