Indian Sign Language Recognition: Database Creation, Hand Tracking and Segmentation

Purva A. Nanivadekar
M.Tech Student, EXTC Department
MPSTME, NMIMS University
Vile Parle, Mumbai, India
Purva.nanivadekar@gmail.com

Dr. Vaishali Kulkarni
Associate Professor, EXTC Department
MPSTME, NMIMS University
Vile Parle, Mumbai, India

***Abstract*—In this paper a Sign Language Recognition system has been proposed. The first step of this system is to create a database of Indian Sign Language. This is done by acquiring the videos from the signers while they are performing the hand gestures. Next step is Hand tracking and Segmentation. This is performed in order to extract features from a particular gesture. A three step algorithm has been used in the proposed system to get better quality hand tracking and segmentation. This algorithm works on motion tracking, edge detection and skin color detection. The system is implemented successfully and results are presented in this paper. The results demonstrate working of motion tracking, edge detection and skin color detection individually as well as their combined effect.**

Keywords—Hand Tracking; Motion tracking; Skin Color Model; Edge detection; Segmentation

I. INTRODUCTION

A Sign Language (SL) is the natural way of communication of the challenged people with speaking and hearing disability. According to the survey by taken by government of India, it is reported that over 21 millions of people in India suffer from some form of the disability. Out of this 7.5% people have disability with speech and 5.8% people face hearing disability [1]. It is necessary that this part of our population should be equally treated as the rest of the society. This is achievable only when all these challenged people are given equal opportunities of education. To bridge the gap between the specially restrained community and the healthy people, more communication between them is necessary. However it requires some additional efforts. The dialogue with them requires an aid of Sign Language. Similar to the spoken language, every country has developed its own Sign Language with a number of grammatical differences. The well-known sign languages are [2]:

American Sign Language (ASL)
British Sign Language (BSL)
Spanish Sign Language
Israeli Sign Language
Indian Sign Language (ISL)

More than 1 million of adults and 0.5 million children in India make use of Indian Sign language [3]. However there is a certain limitation for the development of Indian Sign Language (ISL) due to various socio-cultural factors. Dasgupta et al [4] have mentioned some of such reasons in their work. They include:

a) In rural India, deafness is considered as a punishment for sins.
b) Signing was discouraged by people.
c) Until the late 1970's, it was believed that, there is no such language called ISL.
d) India lacks in research of ISL linguistics.
e) Unavailability of well documented and annotated ISL lexicon.
f) Unavailability of any ISL learning tool.
g) Difficulties in getting sign language interpreters.

More focus on the linguistic studies of ISL was given from 1978 onwards. It was accepted nationwide that ISL is a completely natural language with its own grammar, syntax, phonetics and morphology.

In the later decade of 2000, research began on ISL recognition system. ISL recognition system is a way by which the SL can be decoded and interpreted in the local spoken language. Aleem et al [5] had developed a Gesture Recognition system that examined the input gestures for match with a known gesture in the gesture database. A Gestural database contained all the necessary information that was required for pattern matching. A gesture-to-text dictionary and Speech Synthesis Module Converted word or letters obtained after gesture analysis into corresponding sound. Byung - woo min et al [6], presented the visual recognition of static gesture or dynamic gesture, in which recognized hand gestures were obtained from the visual images on a 2D image plane, without any external devices. Gestures were spotted by a task specific state transition based on natural human articulation. Static gestures were recognized using image movements of hand posture, while dynamic gestures were recognized by analyzing their moving trajectories on the Hidden Markov Models (HMMs). A method had been developed by T. Shanableh [7] for recognizing isolated Arabic sign language gestures in a user independent mode. In this method the signers wore gloves to simplify the process of segmenting out the hands of the signer via color segmentation. The effectiveness of the proposed user-independent feature extraction scheme was assessed by two different classification techniques; namely, K-NN and polynomial networks. Many researchers utilized special devices to recognize the Sign Language [8 - 9]. Hand shape and motion are extracted easily and accurately using these devices. However, the devices were expensive and, crucially, they reduced the naturalness of Sign Language communication. There are many more approaches used to perform feature extraction of the ISL recognition process. These include: palm image extraction method [10], camshift algorithm [11], skin color based model like YCbCr [12], Principal Curvature Based Region Detection (PCBR) method [13] and wavelet packet decomposition [14], and by real time hand tracking used in gesture recognition by Feng-Shen Chen et al [15].

This paper introduces a method to perform Hand Tracking and segmentation for an ISL Recognition system. Section II describes the algorithm of the system; section III includes the implementation along with the results of the system. Section IV has conclusions and discussions.

II. HAND TRACKING AND SEGMENTATION FOR ISL RECOGNITION SYSTEM

The developed system works on the previously captured ISL videos. The mute signers make the ISL gestures which are recorded and later on processed. Before proceeding for hand tracking and segmentation, this video has to be converted to the desired format which is suitable for working. Then the three step algorithm has been deployed on this pre-processed video so as to achieve feature extraction.

A. Image Acquisition:

The first module in the proposed system is the image acquisition module. The image acquisition has been achieved using a video recorder device. A digital camera with standard specifications is an ideal image acquisition device. The recorded videos are of resolution 640 x 480 in a VGA mode of video recording.

B. Video Processing:

The recorded videos have to be preprocessed. First the videos are converted into frames. These frames have been named serially as per their sequence. These frames have been called as O_i, where O stands for original frame. And i stands for current count of the frame. They are then cropped in order to select a specific region. The cropped output is the frame on which the algorithm has been employed. This cropped image has been referred to as $C(O_i)$.

C. Hand Tracking:

Hand tracking has been done by first considering the difference of two adjacent frames. If the current frame is the i^{th} frame, F_i, then the difference image D_i is the difference between i^{th} and $(i+1)^{th}$ frame, $F_{(i+1)}$. The obtained difference image will subtract the background. Only the moving part in both images will be retrieved and all the stationary part will be subtracted.

Along with this, for hand tracking the YCbCr based skin color model has also been employed. Skin color segmentation is performed in YCbCr color space since it reduces the effect of uneven illumination in an image. YCbCr is an encoded nonlinear RGB signal with simple transformation; explicit separation of luminance and chrominance components makes this color space attractive for skin color modeling. RGB color frames $I(m,n,p)$ (where m, n and p are number of rows , number of columns and number of color planes) are converted into YCbCr images using Eq. (1)[16].

$$Y = 0.999R + 0.587G + 0.114B \quad (1)$$
$$Cr = R - Y \qquad Cb = B - Y$$

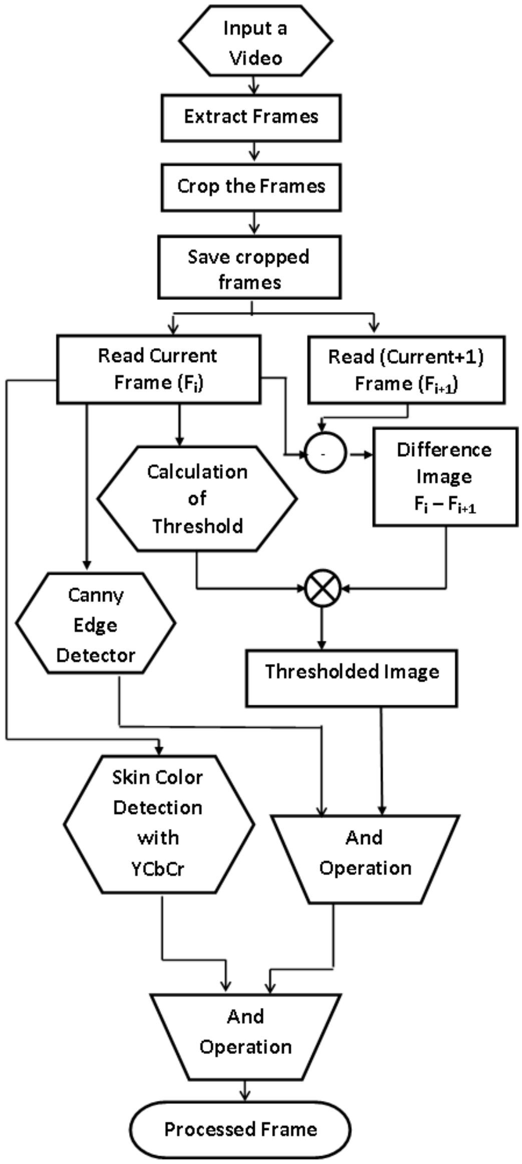


Fig. 1: Flowchart for processing of the frames

YCbCr was developed as part of the ITU-R Recommendation B.T. 601 for digital video standards and television transmissions. It is a scaled and offset version of the Y UV color space. In YCbCr, the RGB components are separated into luminance (*Y*), chrominance blue (*Cb*) and chrominance red (*Cr*). The Y component has 220 levels ranging from 16 to 235, while the Cb, Cr components have 225 levels ranging from 16 to 240. In contrast to RGB, the YCbCr color space is lumaindependent, resulting in a better performance. The corresponding skin cluster is by Eq. (2) [16]:

$$Y > 80 \quad (2)$$
$$85 < Cb < 135 \qquad 135 < Cr < 180$$

Where Y,Cb,Cr = [0,255] .

Chai et al [17] have developed an algorithm that exploits the spatial characteristics of human skin color. A skin color map is derived and used on the chrominance components of the input image to detect pixels that appear to be skin. The algorithm then employs a set of regularization processes to reinforce those regions of skin – color pixels that are more likely to belong to the facial regions. Working in the YCbCr space they have found that the range of Cb and Cr most representatives for the skin – color reference map given by Eq. (3).

$$77 \leq Cb \leq 127 \; and \; 133 \leq Cr \leq 173 \quad (3)$$

D. Hand Segmentation:

After tracking the motion of hand successfully by above two methods, the next step is to obtain the boundaries of the hand movement. This is achieved by performing thresholding and edge detection.

Thresholding is a very important technique for image segmentation. It produces uniform regions based on condition on threshold *T*. *T* can be described as $T=T\{x,y,A(x,y),f(x,y)\}$ where $f(x,y)$ is the gray level of the pixel at (x,y). $A(x,y)$ is the local property of images.

In the proposed system, the thresholding function has been defined as given by Eq. (4),

$$T_i = T\{F_i - F_{(i+1)}\} \quad (4)$$

Where T_i is the threshold image and T is the threshold. F_i and F_{i+1} are the two adjacent frames. The threshold is selected as T= (0.2 x average luminance) of the captured image.

An edge detection algorithm has also been deployed on the original image. A canny edge detector has been used for successful detection of edges. Canny edge detection is based on following criteria:

1. Image smoothening using Gaussian Filter. In this image is blurred in order to remove noise.

2. Finding Gradient: The edges are marked where the gradient s of image has large magnitudes.

3. Non maxima suppression: In this step only local maxima should be marked as edges.

4. Double Thresholding: Here the potential edges are determined using thresholding.

5. Edge tracking by hysteresis: This is the final step in which final edges are determined by suppressing other edges which are not connected to a strong edge.

E. Combined Output:

The final output of the system will be a combined image of segmented output, Edge tracked output and skin color extracted output as given by Eq. (5).

$$C_i = T_i \,\&\, E_i \,\&\, S_i \quad (5)$$

Where C_i: Combined output
T_i: Thresholding output
E_i: Edge detected output
S_i: Skin color model output

III. IMPLEMENTATION

The first step of the system implementation is data collection. As there is no reference video data available for Indian Sign Language, the data has been collected personally at Ali Yavar Jung National Institute of Hearing Handicapped, Bandra, Mumbai. The gestures incorporated in the system are alphabets A to Z, numbers 1 to 10 and few phrases that are be useful in daily conversation and in case of emergency. The data was captured from multiple signers in multiple iterations. The signers were both, males as well as females. To get variations, we introduced one signer who was lefty, and one who had an extra thumb. All the videos were captured in a classroom with various backgrounds like a white board, notice board, a wall etc. All the videos were captured using Nikon Coolpix S9100 in VGA mode with resolution 640 X 480. Fig. 2 show few sample frames for different numbers.

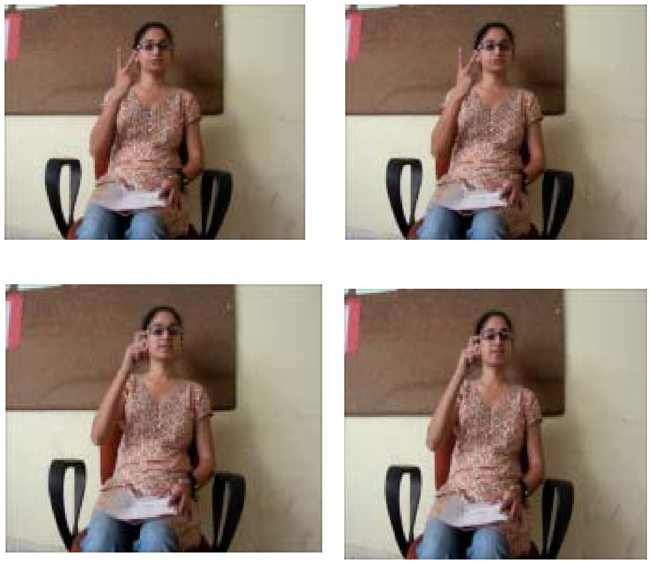


Fig. 2 Sample frames for numbers 2 and 7 of one person

These frames are converted to gray scale images. The frames are then cropped so as to limit the area of working to a specific region. Then two leveled of cropping was performed. In first level the signer was concentrated where as in the second stage only the gesture was cropped. Fig. 3 shows the output of two level image cropping.

This cropped image is used for further processing. The first step of algorithm is to find the difference image as compared to the current and the next frame. Fig. 4 shows the output of the difference image.

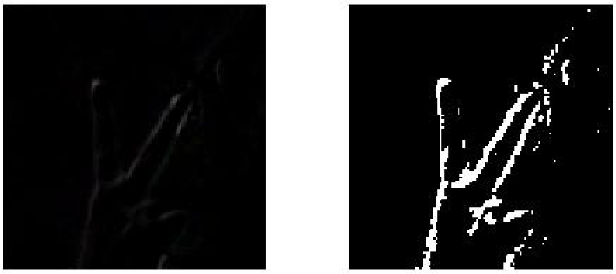
(a) (b) (c)

Fig. 3: Two Level cropping. (a): Original Frame, (b): Gray converted 1st level of cropping, (c): 2nd level of cropping

(a) (b) (c)

Fig. 4: Output of Difference Algorithm. (a): Current Frame, (b): Next Frame, (c): Difference Image.

After obtaining the difference image, segmentation algorithm was applied to the output. Segmentation is achieved using thresholding. In order to achieve optimum thresholding, multiple trials of different threshold values were taken. The best results were obtained with threshold value T to be equal to 0.2 times of the average luminance of original image. Fig. 5 illuminates the output after thresholding.

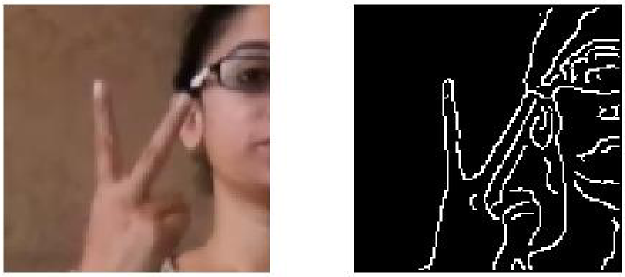


(a) (b)

Fig. 5: Output of Segmentation. (a): Difference Image, (b): Segmented Image

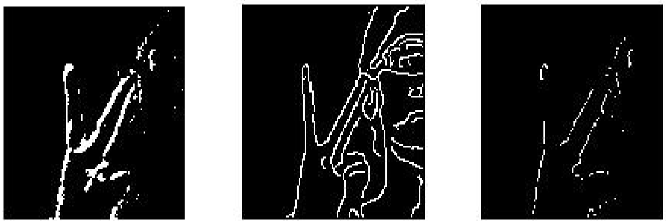
Simultaneously also the Edges of the original image were found. As discussed in section II, canny edge detector was employed to extract edges of the original image. Fig. 6 shows the output of edge detected frame.

Then the output of segmented image and the edge detected image was combined. The combination of these two outputs is obtained by simply using logic and. In Fig. 7 the combined output can be seen.



(a) (b)

Fig. 6: Edge Detection output. (a): Original image, (b): Edge Detected output



(a) (b) (c)

Fig. 7: Combined Output of Segmentation and Edge Detection. (a): Output of Segmentation, (b): Output of Canny Edge Detector, (c): Combined output of Segmentation and edge detection.

In order to receive even better feature extraction, a skin color based model is incorporated in the system. Using YCbCr model, the skin color in the frame is detected. To achieve this, the original image has to be used in RGB domain which then we convert from RGB to YCbCr domain. The skin color model was applied in which the Cb and Cr values of the image are bounded in the specified limits. The image is then converted back to RGB domain and later in gray scale. Fig. 8 shows the output of skin color detection in gray scale.

(a) (b)

Fig. 8: Output of Skin color Model. (a): Original RGB Image, (b): Skin color detected output.

(a) (b) (c) (d)

Fig. 9: Combination of all three detection algorithms. (a): Segmented output, (b): Edge Detection output, (c): Skin color model output, (d): Combined final output.

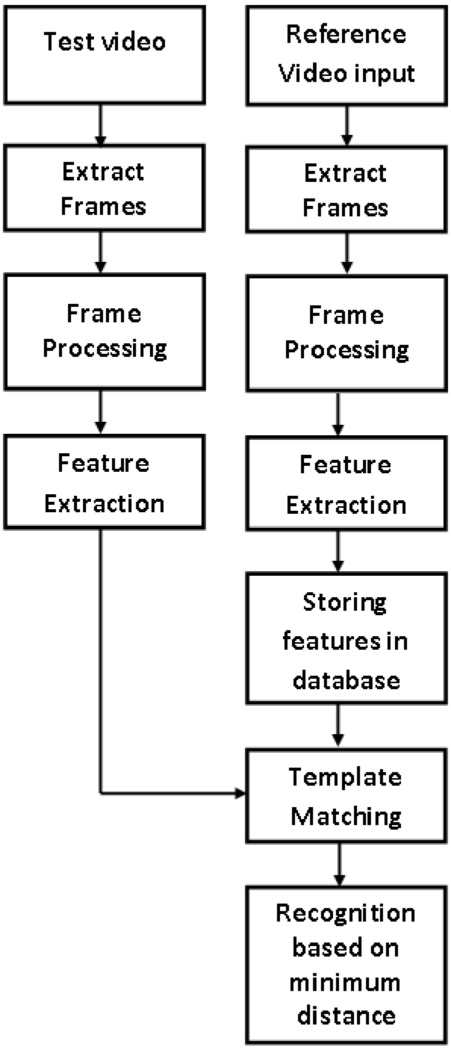


Fig. 10: Flowchart of a complete ISL Recognition system

The final step of the hand tracking and segmentation is combination of all the three images; i.e. segmented image, edge detected image and skin color detected image. Fig. 9 is the output of combination of all three images.

The complete recognition system will follow the sequence given as given below in Fig. 10

IV. CONCLUSION AND FUTURE SCOPE

In this paper Indian Sign Language Recognition is described. The system begins with an Image Acquisition. Hand tracking and segmentation is achieved using proposed algorithm. This algorithm helps in feature extraction of the gestures. The algorithm takes into consideration a motion tracking by implementing a difference image algorithm, pattern recognition by employing edge detection algorithm and hand tracking by employing skin color based model. The advantage of the system inclusion of dynamic gestures.

The next step will be to implement this algorithm to phrases like “Thank you”, “Sorry”, “Help Me”, “Danger” etc. Also we can include the phrases that involve facial movements in the gesture.

ACKNOWLEDGMENT

The authors are highly thankful to “Ali Yavar Jung National Institute for Hearing Handicapped, Mumbai” for their support. Thanks to “Ms. Kinjal Shah, In-Charge, ISL cell” for scheduling the database collection with the signers and helping us to understand basic communication with the mute signers. A special thanks to all the signers who allowed us to capture them and use the recorded data for our project.

REFERENCES

[1] Website:http://censusindia.gov.in/Census_And_You/disabled_population.aspx.

[2] Futane P, Dharaskar R, “Hasta Mudra- An Interpretation of Indian Sign Hand Gestures”, 3rd International conference on Electronics Computer technology, Volume 2, 2011.

[3] Zeshan U., “Indo-Pakistani Sign Language Grammar: A Typological Outline”, Sign Language Studies - Volume 3, Number 2, pp. 157-212,2003.

[4] Dasgupta T., Shukla S., Kumar S., Diwakar S., and Basu A., “A Multilingual Multimedia Indian Sign Language Dictionary Tool”,6th Workshop on Asian Languae Resources, pp. 57-64,2008.

[5] Aleem Khalid Alvi, M. Yousuf Bin Azhar, Mehmood Usman, Suleman Mumtaz, Sameer Rafiq, Razi Ur Rehman, Israr Ahmed T , “Pakistan Sign Language Recognition Using Statistical Template Matching” ,World Academy of Science, Engineering and Technology, 2005.

[6] Byung-woo min, Ho-sub yoon, Jung soh, Takeshi ohashi and Toshiaki jima,” Visual Recognition of Static/Dynamic Gesture: Gesture-Driven Editing System”, Journal of Visual Languages & Computing Volume10,Issue3, pp 291-309,1999.

[7] T. Shanableh, K. Assaleh, “Arabic sign language recognition in user independent mode”, International Conference on Intelligent and Advanced Systems, pp. 597-600, 2007.

[8] C. Vogler and D. Metaxas, “A Framework for Recognizing the Simultaneous Aspects of American Sign Language,” Computer Vision and Image Understanding, vol. 81, no. 3, pp. 358-384, 2001.

[9] W. Gao, G. Fang, D. Zhao, and Y. Chen, Transition Movement Models for Large Vocabulary Continuous Sign Language Recognition”, International Conference on Automatic Face and Gesture Recognition, pp. 553-558, 2004.

[10] P. Subha Rajam, G. Balakrishnan, “Real Time Indian Sign Language Recognition System to aid Deaf-dumb People”, 13th International conference on Communication technology, pp. 737-742, 2011.

[11] A. Ghotkar, R. Khatal, S, Khupase, S. Astani, & M, Hadap, “Hand Gesture Recognition for Indian Sign Language”, International

Conference on Computer Communication and Informatics, pp. 1-4, 2012.

[12] J. Rekha, J. Bhattacharya, S. Majumder, “Shape, Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognition”, 3rd International conference on Trendz in Information Sciences and Computing, pp. 30-35, 2011.

[13] H. Deng, W. Zhang, E. Mortensen, T. Dietterich & L. Shapiro, "Principal Curvature- Based Region Detector for Object Recognition", International Conference on Computer Vision and Pattern Recognition pp. 1-8, 2007.

[14] M.C. Padma & P.A. vijaya, "Global Approach For Script Identification Using Wavelet Packet Based Features", International Journal of Signal Processing, Image Processing And Pattern Recognition, Vol.3, pp. 29-40, 2010.

[15] F.S. Chen, C. Fu, & C. Huang, “Hand gesture recognition using a real-time tracking method and hidden Markov models”, ELSEVIER, Image and Vision Computing, Volume 21, Issue 8, pp. 745–758, 2003.

[16] T. Mahmoud, “A New Fast Skin Color Detection Technique”, World Academy of Science, Engineering and Technology, pp.498-502, 2008.

[17] D. Chai, and K.N. Ngan, "Face segmentation using skin-color map in videophone applications". IEEE Trans. on Circuits and Systems for Video Technology,Volume 9, Issue 4,pp. 551-564, 1999.