ISSN:0975-9646

Ashmi Chheda et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (2) , 2016, 562-565

A Study on Vision Based Sign Language Translation

Ashmi Chheda[1], Janvi Beed[2], Siddhi Barbhaya[3] Prof. Bhakti Palkar

Department of Computer Engineering
[1,2,3] K. J. Somaiya College of Engineering, Vidyavihar, Maharashtra, India, Mumbai-77

Abstract: Computer recognition of sign language is an important research problem for enabling communication with speech and/or hearing impaired people. This report presents a computer based Vision based sign language translation system for automatic translation of Indian Sign Language (ISL) into text. This will assist the hearing/speech impaired people to significantly communicate with all other people using their hand gestures, and will make the communication effective and efficient. It will also eliminate the need of human translator and will enable the hearing and/or speech impaired people to communicate independently. The video is acquired using built in web camera, the images acquired from the video is then pre-processed and then gesture recognition is performed. This system translates one handed sign representations of numbers (0-9) and alphabets (A-Z).

I. Introduction

Sign Language is the way by which people with hearing and speaking disabilities communicate with others. Although it is very easy to communicate amongst themselves, it becomes a challenge for them to communicate with the normal people. [6] In writing, it was found that include of listening to debilitated individuals India, is more contrasted with different nations. Not every one of them use ISL but rather, more than one million hard of hearing grown-ups and around half million hard of hearing kids use ISL as a method of correspondence. Hard of hearing individuals, who live in towns for the most part, don't have admittance to communication via gestures. Be that as it may, in every vast town and urban areas over the Indian subcontinent, hard of hearing individuals use communication through signing which is not standard gesture based communication. Vision based Sign Language Translation will be a product that will conquer any hindrance between quiet individuals and ordinary individuals with the goal that typical individuals can comprehend the gesture based communication utilized by quiet individuals. Vision based Sign Language Translation is a use of Digital Image Processing. It is a desktop framework which is utilized to perceive and interpret nonstop Indian gesture based communication to English content. This framework deciphers letter sets (A-Z), numbers (0-9) and signals. This framework is mechanized, and with an easy to use interface. This will be finished by taking a video of a sentence in gesture based communication and giving its resultant significance as content.

Vision based approach have been discussed for interpreting the Indian sign language using hand modality. A Typical Hand Gesture Recognition system consists of mainly four modules: Gesture acquisition, Tracking and segmentation, Feature extraction and description, Classification and recognition. This paper focuses on a study of sign language interpretation system with reference to vision based hand gesture recognition. An attempt has also been made to explore about the need and motivation for interpreting ISL, which will provide opportunities for hearing impaired peoples in Industry Jobs, IT sector Jobs, and Government Jobs.

The intended user group will be:
1) Mute people (Will perform sign language in front of a web camera);
2) Normal people (who will get the output of the sign language in the form of text);
3) There won't be much dependency or the presence needed for a mediator to translate for the dumb person. Our system does it all;
4) To provide a new path of communication experience between a dumb and normal person with a user friendly interface.

II. Methodology

This paper presents a real-time vision-based system for recognizing Indian Sign Language (ISL) using a camera. This system is broken down into three main parts viz. image acquisition, followed by image processing for extracting the features for gesture recognition and last comes the recognition stage where signs are identified and text is given as output. This part explains the details of these processes. The program starts by capturing a video, and then the video is broken down into images based on frames. The acquired images are then pre-processed to remove the noise. The images are processed to identify the region of interest. The unique features of each sign in the region of interest are extracted to be used in the recognition stage. In the recognition stage, the features extracted are compared with the available database of pattern matching templates.

2.1 Image Acquisition

A video is captured at around 30 frames per second. [1] This number of frames per second is enough for computation. More number of frames will lead to more time for computation as more data needs to be processed.

2.2 Environmental Setup

The image acquisition process is subjected to many environmental factors like the exposure of light, the background and foreground objects. In order to make feature extraction easier later on, a plain background (white) is easier to work on.

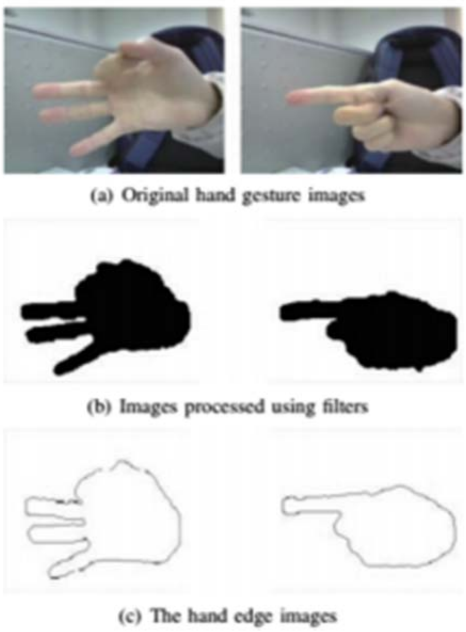
2.3 IMAGE PRE-PROCESSING

This part consists of hand segmentation followed by morphological operations. One method for this is proposed an adaptive skin color model for hand segmentation by mapping YCbCr color space into YCbCr color plane. [2] Another method for hand segmentation and tracking is based on HSV histogram. [3] These methods can segment the hand in simple as well as complex background. For simplicity purposes, the hand segmentation is performed by transforming the image acquired into black and white image, where the background will be white and the foreground i.e. the hand will be black. Another way to preprocess the images acquired is the color thresholding method. [1] Using this method, the color region can be segmented and the position of the region of interest can be determined. Then, the color segmented images are analyzed to obtain the unique features of each sign in the sign language. But this method is suitable only for alphabets' (A – Z) and numbers' (0 – 9) recognition, because it uses the unique combination of fingertip position for feature extraction of each sign. [8] The Gauss-Laplace edge detection method can also be used to get the hand edge. The experimental result shows the Gauss-Laplace algorithm is used to effectively implement the hand edge detection. The Gauss-Laplace operator used is template the figure 1.

$$\begin{bmatrix} -2 & -4 & -4 & -4 & -2 \\ -4 & 0 & 8 & 0 & -4 \\ -4 & 8 & 24 & 8 & -4 \\ -4 & 0 & 8 & 0 & -4 \\ -2 & -4 & -4 & -4 & -2 \end{bmatrix}$$

Figure 1

The hand edges are obtained when the hand images are processed by using the Gauss-Laplace algorithm, and Figure 2 shows the results.



(a) Original hand gesture images

(b) Images processed using filters

(c) The hand edge images

Figure 2

An alternate method to pre-process an RGB Image is by converting it to binary image and then performing boundary extraction using the morphological operation of erosion. It can be demonstrated in the following figure 3.

Figure 3

2.4 SIGN RECOGNITION

The final stage of this process is to identify the sign and produce textual output of its meaning. There are multiple ways to do this too, although the procedure remains the same. The technique to be chosen to recognize the sign depends on the method used for preprocessing the image. If color thresholding followed by fingertip position extraction is used, then the recognition will be based on the position of finger in the bounding box. [1] A threshold value is set for the difference between the database image and the input image. If the difference obtained is below this threshold value, a match is said to have been found.

Another technique that can be used is finger counting logic. [4] In this technique, active fingers involved in input gestures are calculated using variable Centroid distance feature. Then maximum Euclidian distance 'Dmax' between two points, Centroid and the point located on counter of edge image is calculated.

Another technique is by using the pattern matching algorithm. If the preprocessed image is a black and white image, then the database needs to be built using black and white images only. In this method, the input images will be pattern matched with database images. The one with the nearest pattern to the input will be declared as the output. One of the best algorithms available for pattern matching of images is SIFT (Scale Invariant Feature Transform) which is used to detect and describe local features in images. [5] SIFT is preferable algorithm due to advantages like good recall rates (accuracy), features are robust to occlusion and clutter and comparatively efficient. A brief description of the SIFT image feature matching steps is carried out as follows:

(1) The first step is to detect scale space extreme points
   i. Establishment of Gaussian pyramid.
   ii. Generation of DOG (difference of Gaussian) pyramid.
   iii. Extreme point detection in DOG space.

(2) The next step is to precisely position the key points.

(3) Then, determination of the main direction of the key points has to be done.

(4) After that, generation of the SIFT feature vector (Generate descriptors of key points) has to be done.

(5) Matching. After generating the feature vector of the two images, regard the feature vector's Euclidean distance of the key points as the similarity metric of key points in the two images.

Another best algorithm for pattern matching we have studied is SURF Algorithm (Speeded-Up Robust Feature). [9] SURF is proposed as a speed-up version of SIFT [5] and is mainly divided into three parts: 1) The key points (feature points) extraction. Using Hessian matrix to detect the extreme point in Difference of Gauss (DOG) [6] and take it as the key point of potential, eliminating the unstable points; 2) The key points local feature description. Calculating the Haar-wavelet responses of x and y direction in a small area around the key point to obtain the direction and the description of feature point; 3) The key points matching. When the description vectors of the key points in two images are the same, the first two points matched successfully are the matching points.

Researches have shown that SURF is faster than SIFT but it is unstable. The two algorithms have been studied for the following three scenarios:

1) Effect of the number of the n-nearest key points related with the noise.
2) Robustness of the local feature descriptor.
3) Performance evaluation (Computation Time)

Following are the results obtained for the above three scenarios:

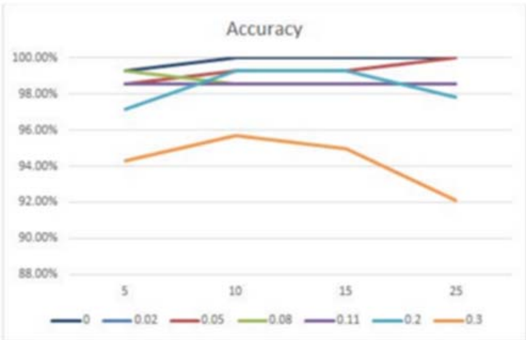


Figure 4. Result of 1st scenario for SIFT.

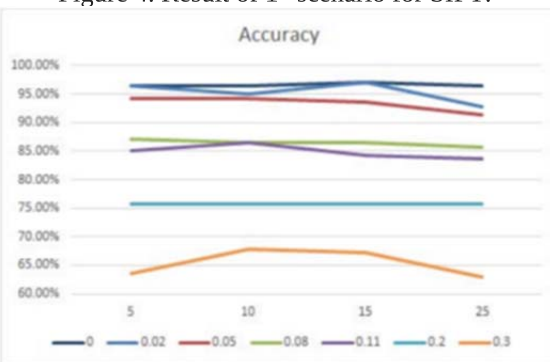


Figure 5. Result of 1st scenario for SURF.

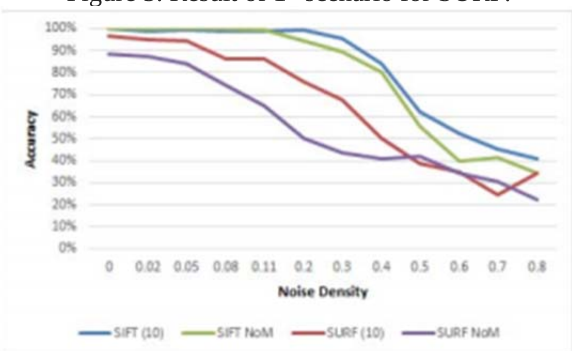


Figure 6. Result of 2nd Scenario.

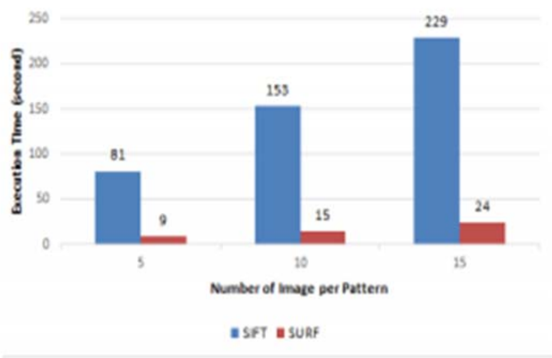


Figure 7. Result of 3rd scenario.

From the above results of the graph, it can be interpreted that the local feature descriptor performed better in SIFT than SURF. But when execution time is vital, SURF is preferable over SIFT.

2.5 Database Creation

A very straightforward method to store the initial images for comparison with the images obtained from input video would be making use of a folder. This would mean storing images in a folder and accessing the folder location to retrieve each image one by one for comparing. It is a very easy task in MATLAB. But at the same time is also quite redundant and time consuming.

To provide efficiency, database creation for storing information for pattern matching can be done by installing ODBC driver using Database Explorer App that is available in MATLAB software. By installing ODBC driver, direct usage of the database can be done. The other option is connecting to an external database and commands can be used for extracting those images and it can be then used for matching the images.

III. Result

This sign language translator will be able to translate numbers (0-9) and alphabets (A-Z). All the signs can be translated real-time. But signs that are similar in gesture and posture to another sign can be misinterpreted, resulting in a decrease in accuracy of the system.

IV. Conclusion

In this paper, various algorithms and methods by which the process of sign language translation can be performed are discussed. The advantages of some methods over others are also discussed. The vision of an efficient system to translate sign language to text is quite achievable, but the challenges lie in optimization.

References

[1] Madhuri, Y.; Anitha, G.; Anburajan, M., "Vision-based sign language translation device," in *Information Communication and Embedded Systems (ICICES), 2013 International Conference on* , vol., no., pp.565-568, 21-22 Feb. 2013doi: 10.1109/ICICES.2013.6508395

[2] Dawod A.Y., Abdullah 1. and Alam MJ., "Adaptive skin color model for hand segmentation," Computer Applications and Industrial Electronics (ICCAIE), 2010 International Conference on, vol., no., ppA86,489, 5-8 Dec. 2010.

[3] Dhruva N., Rupanagudi S.R., Sachin S.K., Sthuthi 8., Pavithra R. and Raghavendra, "Novel segmentation algorithm for hand gesture recognition," Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on, vol., no., pp.383,388, 22-23 March 2013.

[4] Bhame, V.; Sreemathy, R.; Dhumal, H., "Vision based hand gesture recognition using eccentric approach for human computer interaction," in Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on , vol., no., pp.949-953, 24-27 Sept. 2014
doi: 10.1109/ICACCI.2014.6968545

[5] Zhou, Qiangqiang; Zhao, Zhenbing, "Substation equipment image recognition based on SIFT feature matching," in Image and Signal Processing (CISP), 2012 5th International Congress on , vol., no., pp.1344-1347, 16-18 Oct. 2012
doi: 10.1109/CISP.2012.6469854

[6] Archana S. Ghotkar, Dr. Gajanan K. Kharate, "Study of vision based hand gesture recognition using Indian sign language ", International journal on smart sensing and intelligent systems vol. 7, no. 1, March 2014

[7] Dominikus Willy, Ary Noviyanto, Aniati Murni Arymurthy, "Evaluation of SIFT and SURF Features in the Songket Recognition", ICACSIS 2013 ISBN: 978-979-1421-19-5

[8] Chenglong Yu, Member, IEEE, Xuan Wang, Member, IEEE, Hejiao Huang, Member, IEEE, Jianping Shen, Kun Wu, "Vision-Based Hand Gesture Recognition Using Combinational Features", 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

[9] Penghui Jiang, Shengjie Zhao, Samuel Cheng, "Rotational Invariant LBP-SURF for Fast and Robust Image Matching", 978-1-4673-8118-5/15/$31.00/©2015 IEEE