IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

An Aid for the Hearing Impaired

Nagarjun T R M[1], Abhilash V Pattan[2], Akshay Kumar H C[3], Apoorva Prabhudeva[4], Veena Karjigi[5]

Department of Electronics and Communication, Siddaganga Institute of Technology, Tumakuru[1, 2, 3, 4, 5]

Abstract: Sign language is a language which is used for communication between the hearing/speech impaired people along with others who understands it. Sign language relies on sign patterns, i.e., body language, orientation and movements of the arm to facilitate understanding between people. This paper is aimed at reducing the communication gap between the normal people and hearing/speech impaired people by converting sign language into English speech or text. The system will convert the data into speech/text. The whole system will be implemented on a MATLAB and Simulink.

Keywords: Sign language, Image Processing, Colour Models, Segmentation, Morphology, Otsu's algorithm, Gaussian PDF, Simulink.

I. INTRODUCTION

Advancements in technology have made our lives very easy. Introduction of automated cars [1], smart trolleys [2], online shopping [3] and many more inventions like these have changed our way of living. It has not only added luxury to our lives, but also has served in protecting us [4], [5]. On the other hand, new innovations have also helped the community of the physically challenged to get connected to the world: right from helping the blind in reading the English literature [6], to accessing ATMs [7], aiding the patients suffering from ALS to commute [8] and communicate [9]. Similarly, a lot of research is being carried out to assist the hearing impaired community as well.

Sign language is a language which uses a collection of hand gestures, symbols and body language to convey meaning. This language is completely dependent on orientation and movement of the hands, arms or body, and also, the facial expressions. The oral language, which we speak today with sounds, was originally evolved from the sign language.

Sign language is the most preferred language when it comes to communicating with the hearing and/or speech impaired. Learning and understanding this mode of communication isn't just for the hearing impaired, but also for those who are interested in conversing with the hearing impaired. Thus, in order to bridge the communication gap between the normal community and the community of the deaf and the dumb, a system that can convert their "gesture" language to English has to be emphasized.

The paper has been divided into the following sections: Section 2 described existing approaches for sign language recognition system. Section 3 deals with proposed approach. Section 4 and Section 5 illustrate the results and conclusion.

II. EXISTING TECHNIQUES

Many attempts for this humanitarian cause have been recorded before. Few of them are described in this section. As explained in [10], the gesture recognizing glove, also terms as data glove, makes use of a low cost packaging material (velostat) for making piezoresistive sensors. These flex sensors detect a bend in fingers and map this data to a character set by implementing a Minimum Mean Square Error machine learning algorithm. The recognized character is transmitted via Bluetooth, to an Android phone, which performs a text to speech conversion. The motivation for Hand Talk is to compare hand configurations with sign language charts and generate artificial speech which articulates the gestured words. The primary input of the system would be the pose and orientation of the hand. The main focus is on acquiring to what extent each of the finger joints is bent. Upon acquiring this data it is encoded and transmitted wireless to a mobile device. The software in the mobile device would guess the shape and orientation of the hand based on the received data.

In [11], a wireless hand gesture recognition glove is proposed for real time translation of Taiwanese sign language. To discriminate between different hand gestures, flex and inertial sensors are embedded into the glove so that the three most important parameters, i.e., the posture of fingers, orientation of the palm, and motion of the hand, defined in Taiwanese Sign Language can be recognized without ambiguity. The finger flexion postures acquired by flex sensors, the palm orientation acquired by G-sensor, and the motion trajectory acquired by gyroscope are used as the input signals of the proposed system. The input signals are acquired and examined periodically to see if it is a legal sign language gesture or not. Once the sampled signal can last longer than a predefined clock cycles, it is regarded as a valid gesture and will be sent to cell phone via Bluetooth for gesture discrimination and speech translation. These hand gestures correspond to letters, and in order to form a word many gestures have to be shown.

The authors in [12], propose an approach for hand gesture recognition based on detection of shape based features. The setup consists of a single camera to capture the gesture formed by the user. Segmentation is performed on the image acquired. Next, orientation detection, feature extraction and classification are performed. Here, words

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

are identified effortlessly, but forming sentences has not been proposed.

Since, very less research has been done on identification of sentences, we propose a novel method to recognize the gestures shown.

III. PROPOSED ALGORITHM

A. Setup

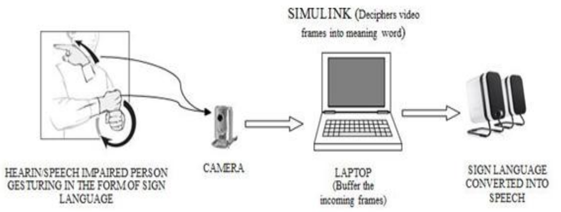


Fig 1: System Overview of Sign language conversion system

The setup is shown in Fig 1. The physically disabled person gestures the sign language to the camera placed in front of him at a distance of 150 cm. The camera captures the gestures at 30 frames per second. The minimum resolution of the camera must be 2MP. The video frames are sent to the processor, in our case Simulink), which analyses the gestures and converts into speech and text.

B. Algorithm

The flow of the proposed algorithm is shown in Fig 2.

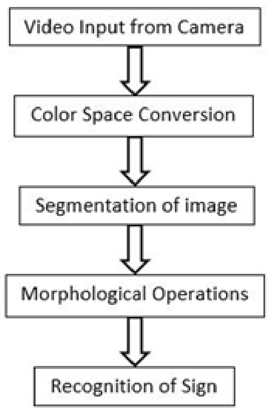


Fig 2: Flowchart of proposed algorithm

1) Image acquisition and Preprocessing

The first stage of any computer vision system is the Image Acquisition stage. Here, the images of the gestures shown are acquired and sent to the processor. The acquired image is in RGB color model. Since this color model depends on the brightness in the environment, working directly on the images procured might prove to be risky. Thus, the image is converted to the Chroma Red and Chroma Blue Components of the YCbCr model, to remove the brightness factor. The conversion is made using the formulae as mentioned in equation (1) and equation (2) respectively [13].

$$\mathrm{Cb} = 0.148\mathrm{R} + 0.291\mathrm{G} + 0.439\mathrm{B} + 128 \quad (1)$$

$$\mathrm{Cr} = 0.439\mathrm{R} + 0.38\mathrm{G} + 0.071\mathrm{B} + 128 \quad (2)$$

The skin is clearly distinguishable from the background in the Cr component than in Cb component. This feature was evidently noticed for a dataset of 96 images with varying backgrounds.

2) Segmentation

The next step is to perform segmentation to extraction of the gesture shown. Initially, this process was conducted using Otsu's algorithm [14] of thresholding. This method gives a threshold after processing the pixel intensity values of the grayscale image. The algorithm is described as follows.

Step1: Acquire RGB images
Step2: Convert the image to Cr color plane
Step3: Perform histogram on the image, where the bins are on the x-axis and the counts on the y-axis.
Step4: Find the sum of all counts
Step5: Normalize each count and store it in an array sumN
Step6: Find cumulative sum of sumN and store n sumNC
Step7: Multiply the bins with sumNC. The product obtained would be EM.
Step8: Find the cumulative sum of EM and save it as sumEM
Step9: Find the final value of sumEM and store as sumEMf
Step10: Find the variance of the image using the equation (3).

$$\sum \mathrm{b}^2(\mathrm{i}) = \frac{(\mathrm{sumEMf} * \mathrm{sumNC}\ (\mathrm{i}) - \mathrm{sumEM}\ (\mathrm{i}))^2}{\mathrm{sumEM}\ (\mathrm{i}) * (1 - \mathrm{sumEM}\ (\mathrm{i}))} \quad (3)$$

where, i ranges from 1 to number of bins.

Step11: Find the maximum value of variance obtained (maxv)
Step12: If max v is finite, then find the index value of maxv from variance array obtained and store it in variable indexv.

$$\text{Step13: Threshold} = \begin{cases} 0, \text{if maxv is infinite} \\ \frac{\text{indexv} - 1}{\text{number of bins} - 1}, \text{ otherwise} \end{cases} \quad (4)$$

The obtained threshold has to be compared with the pixel intensity values of the image. If the threshold is smaller than the pixel value, then the same pixel has to be converted to white, else change the pixel to black.

The Otsu's algorithm did not yield the desired output. Thus, elliptical Gaussian joint probability density function (pdf) has been implemented for improved results. The skin color distribution is modeled by using Gaussian pdf, which is defined as:

$$\mathrm{p(c|skin)} = \frac{1}{2\prod|\sum s|^{\frac{1}{2}}} e - \frac{1}{2}(\mathrm{c} - \mu \mathrm{s})\mathrm{T} \sum_s^{-1}(\mathrm{c} - \mu\mathrm{s}) \quad (5)$$

Here, c is a color vector and μ_s and $\sum_s$ are the distribution parameters (mean vector and covariance matrix respectively). The model parameters are estimated from the training data by equations (6) and (7).

$$\mu_s = \frac{1}{n}\sum_{j=1}^{n} c_j \quad (6)$$

$$\sum s = \frac{1}{n-1}\sum_{j=1}^{n}(cj - \mu_s)(c\text{-}\ \mu_s)^T \quad (7)$$

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

where n is the total number of skin colour samples c_j . The p(c|skin) probability can be used directly as the measure of how "skin-like" the c color is or alternatively, the Mahalanobis distance from the c color vector to mean vector μ_s, given the covariance matrix Σ_s can serve for the same purpose, which is defined as:

$$\lambda_s = (c - \mu_s)^T \Sigma_s^{-1}(c - \mu s) \qquad (8)$$

The single Gaussian model is build by taking a 320x320 pixel image shown in Figure. 3 with 256 skin samples of 20x20 pixel size each. The two parameters mean (μ_s) & covariance (Σ_s) are obtained from the skin sample image. The Fig. 3 is converted into YCbCr color model and the skin samples are mapped as shown in Fig 4 plotting Cb vs. Cr. It is observed that the unnecessary intensity are removed and mapped for only the skin intensity. Fig. 5 shows the Mesh plot of Cb and Cr with respect to its count.

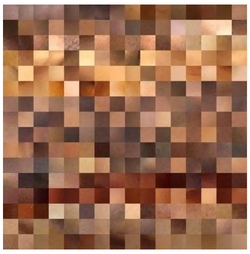


Fig 3: Skin samples image

3) Morphological operations
When segmentation is performed on the video frame, along with gestures, noise pixels might appear as well. Thus, morphological operations such as erosion and dilation have to be performed to obtain only the whit pixels corresponding to the gestures.

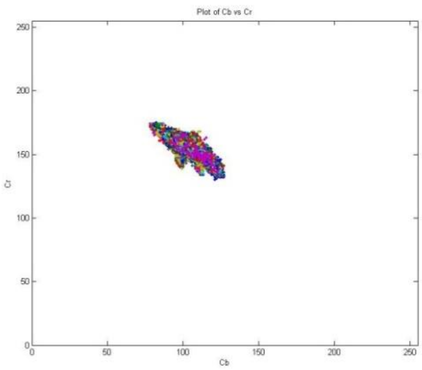


Fig 4: Plot vs. Cb and Cr

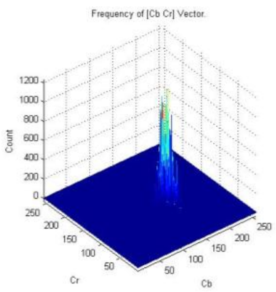


Fig 5: 3D plot of Cb, Cr and Count

4) Recognition of gestures
After performing Morphological operations, the image is divided into equal number of square shaped blocks along rows and columns, and it is further divided into sub-blocks in order to accurately measure the co-ordinate of any sub-block. A square shaped window is defined for every sign, shown in different regions and the co-ordinates of the sub-block included within the window along with a range of number of white pixels included within that particular region is noted down. This is repeated for all the regions were sign can appear. The same is illustrated in Fig 6.

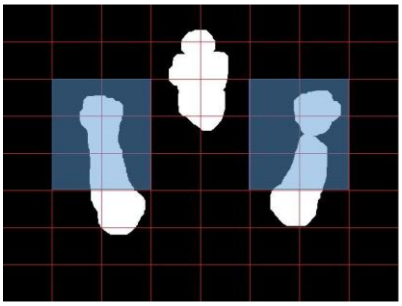
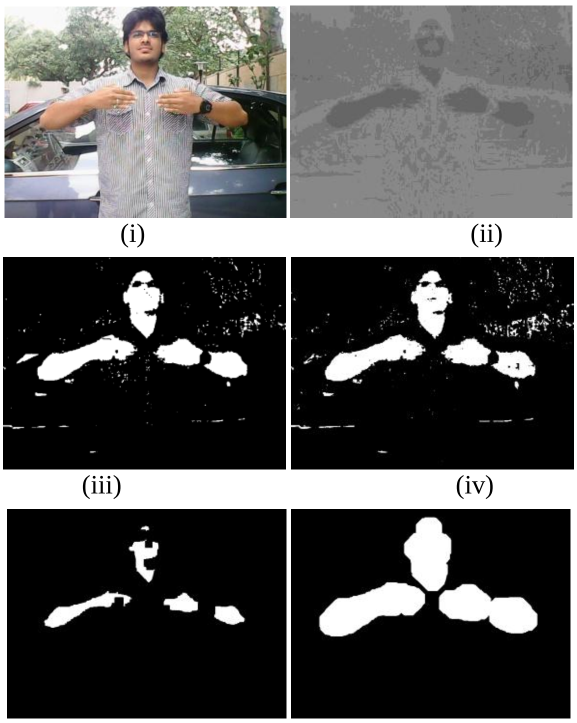


Fig 6: Division of image into blocks and assigning a window for sign

When the system is turned on, it scans and counts the number of white pixels only for the region of co-ordinates defined earlier and if a sign appears within a particular region with number of white pixels within the defined range, corresponding label is assigned to that sign.

IV. RESULTS

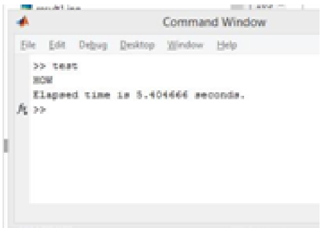
An image frame showing "HOW" has been captured and the processing outputs at each stage have been showcased in Fig 7.



(i) (ii) (iii) (iv) (v) (vi)

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016



(viii)

Fig 7: Detection of sign "HOW": (i) RGB image (ii) Cr component (iii) Segmentation based on Otsu's algorithm (iv) Segmentation based on Gaussian pdf (v) Erosion output (vi) Dilation result (vii) MATLAB showing the gesture in text form

The time consumed by both the segmentation methods is recorded. Otsu's algorithm needs 1.7 seconds, while pdf needs 5.4 seconds. The time required to recognize the sign using Otsu algorithm is much less compared to the Gaussian method. Hence, the system would be faster when segmentation is performed using Otsu algorithm. But, the disadvantage of Otsu algorithm is that it produces single or narrow threshold value and pixels which are also considered as skin might be below this threshold value and hence results in a lot of noise when thresholding is performed. In case of segmentation using Gaussian based modeling, the results are more accurate but the time required in recognizing the gesture is very large and hence, it would not be suitable for real time application.

V. CONCLUSION

As accuracy has been stressed upon, Gaussian pdf method is used for identifying the gestures. Using the algorithm flow explained in Section 3, 20 sentences were identified, which used commonly used words such as "HOW", "ARE", "YOU", "MOM", "DAD", "WHERE", "HOW", "DOING", "HELLO" and "WHAT". Hence, an approach to reduce the communication gap between the normal people and disabled people by converting sign language into English speech or text.

ACKNOWLEDGMENT

Our sincere thanks to **Mr. Sudhir Rao Rupanagudi** from World Serve Education, for contributing towards development of this work.

REFERENCES

[1] C. R. Prashanth, T. Sagar, N. Bhat, D. Naveen, S. R. Rupanagudi and R. A. Kumar, "Obstacle detection & elimination of shadows for an image processing based automated vehicle," Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on, Mysore, 2013, pp. 367-372.
[2] S. R. Rupanagudi et al., "A novel video processing based cost effective smart trolley system for supermarkets using FPGA, "Communication, Information & Computing Technology (ICCICT), 2015 International Conference on, Mumbai, 2015, pp. 1-6.
[3] S. u. Rehman and J. Coughlan, "Building trust for online shopping and their adoption of e-commerce," Information Society (i-Society), 2012 International Conference on, London, 2012, pp. 456-460.
[4] S. R. Rupanagudi, Ranjani B. S., P. Nagaraj, V. G. Bhat and Thippeswamy G, "A novel cloud computing based smart farming system for early detection of borer insects in tomatoes," Communication, Information & Computing Technology (ICCICT), 2015 International Conference on, Mumbai, 2015, pp. 1-6.
[5] S. R. Rupanagudi et al., "A novel video processing based smart helmet for rear vehicle intimation & collision avoidance," 2015 International Conference on Computing and Network Communications (CoCoNet), Trivandrum, 2015, pp. 799-805.
[6] S. Shastry, G. Gunasheela, T. Dutt, D. S. Vinay and S. R. Rupanagudi, ""i" — A novel algorithm for optical character recognition (OCR)," Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on, Kottayam, 2013, pp. 389-393.
[7] Dhruva N, Rupanagudi SR, et al (2013) Novel Algorithm for Image Processing Based Hand Gesture Recognition and Its Application in Security. Advances in Computing, Communication, and Control. doi: 10.1007/978-3-642-36321-4_51
[8] Rupanagudi SR, Bhat VG et al (2015) Design and Implementation of a Novel Eye Gaze Recognition System Based on Scleral Area for MND Patients Using Video Processing. Advances in Intelligent Informatics. doi:10.1007/978-3-319-11218-3_51
[9] S. R. Rupanagudi, B. S. Ranjani, P. Nagaraj and V. G. Bhat, "A cost effective tomato maturity grading system using image processing for farmers," Contemporary Computing and Informatics (IC3I), 2014 International Conference on, Mysore, 2014, pp. 7-12.
[10] C. Preetham, G. Ramakrishnan, S. Kumar, A. Tamse and N. Krishnapura, "Hand Talk-Implementation of a Gesture Recognizing Glove," India Educators' Conference (TIIEC), 2013 Texas Instruments, Bangalore, 2013, pp. 328-331.
[11] Kau, Lih-Jen; Su, Wan-Lin; Yu, Pei-Ju; Wei, Sin-Jhan, "A real-time portable sign language translation system," in Circuits and Systems (MWSCAS), 2015 IEEE 58th Inter-national Midwest Symposium, pp.1-4, 2-5 Aug. 2015
[12] Panwar, M.; Mehra, P.S., "Hand gesture recognition for human computer inter-action",2011 International Conference on Image Information Processing (ICIIP), pp.1-7,IEEE 2011
[13] S. R. Rupanagudi, Vikas N S, V. C. Bharadwaj, Manju, Dhruva N and Sowmya K. S., "Novel methodology for blink recognition using video oculography for communicating," Advances in Electrical Engineering (ICAEE), 2014 International Conference on, Vellore, 2014, pp. 1-6.
[14] NOBUYUKI OTSU,"A Threshold Selection Method from Gray-Level Histograms," in Systems, Man and Cybernetics, IEEE Transactions, pp.62-66, Jan. 1979