Real Time Indian Sign Language Recognition System to aid Deaf-dumb People

P. Subha Rajam
Assistant Professor [SE.G]/ IT Department
J.J. College of Engineering & Technology,
Trichy, Tamilnadu, India.
subha8892@yahoo.co.in

Dr. G. Balakrishnan
Director,
Indra Ganesan College of Engineering,
Trichy,Tamilnadu, India.
balakrishnan.g@gmail.com

Abstract— **The Sign Language is a method of communication for deaf – dumb people. This paper proposes a method that provides a basis for the development of Sign Language Recognition system for one of the south Indian languages. In the proposed method, a set of 32 signs, each representing the binary 'UP' & 'DOWN' positions of the five fingers is defined. The images are of the palm side of right hand and are loaded at runtime i.e. dynamic loading. The method has been developed with respect to single user both in training and testing phase. The static images have been pre-processed using feature point extraction method and are trained with 10 numbers of images for each sign. The images are converted into text by identifying the finger tip position of static images using image processing techniques. The proposed method is able to identify the images of the signer which are captured dynamically during testing phase. The results with test images are presented, which show that the proposed Sign Language Recognition System is able to recognize images with 98.125% accuracy when trained with 320 images and tested with 160 images.**

Keywords— **Indian Sign Language, Image processing, Pattern recognition.**

I. INTRODUCTION

The sign language is an important method of communication for deaf-dumb persons. As sign language is well structured code gesture, each gesture has a meaning assigned to it. It can be used to express complex meanings by combining basic elements. In the last several years there has been an increased interest among the researchers in the field of sign language recognition to introduce means of interaction from human – human to human – computer interaction.

Deaf and Dumb people rely on sign language interpreters for communications. However, finding experienced and qualified interpreters for their day to day affairs throughout life period is a very difficult task and also unaffordable. Hence, human – computer interaction system will prove to be a reliable and consistent solution to such persons. There has been a widespread research in this area being carried out over two decades. Earlier researchers [1] mostly focused on the capture and classification of the gestures of sign language. There were contributions for American [4], Australian [5], Korean [6], and Japanese [7] sign languages. Since then many techniques and algorithms have been proposed using a variety of methods based on sensor fusion, signal processing, image processing, and pattern recognition methods. The application was extended to several international sign languages including Chinese [8], and Arabic [9] [10]. There have been no such notable contributions for South Indian Languages by any of the researchers in this area. There may be different regional versions available in a particular language. However, the sign language is common and applicable to any variant of language. This paper deals with a system which recognizes the Indian Sign Language for human – computer interaction. To help people with such disabilities, the proposed approach is able to formulate a set of 32 combinations of binary number of signs, each representing the dynamic images of the palm side of right hand.

978-1-61284-307-0/11/$26.00 ©2011 IEEE

II. EARLIER WORKS

Aleem et al, had developed a Gesture Recognition systems examined the input gestures for match with a known gesture in the gesture database. Gesture Database Contained the necessary information required for pattern matching as well as a gesture-to-text dictionary and Speech Synthesis Module Converted word or letters obtained after gesture analysis into corresponding sound [1].

Byung - woo min et al, presented the visual recognition of static gesture or dynamic gesture, in which recognized hand gestures 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 moments of hand posture, while dynamic gestures were recognized by analysing their moving trajectories on the Hidden Markov Models (HMMs) [2].

A method had been developed by T. Shanableh 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 [3]. Many researchers utilized special devices to recognize the Sign Language [12][13]. Hand shape and motion are extracted easily and accurately using these devices. However, devices were expensive and, crucially, they reduced the naturalness of Sign Language communication.

III. PROPOSED METHODOLOGY

In the proposed method, 32 combinations of binary number sign are developed by using right hand palm image, which are loaded at runtime. An image captured at run time is scanned to identify finger tip positions of the five fingers namely little fingers, ring finger, middle, index finger and thumb finger as shown in Fig 1.

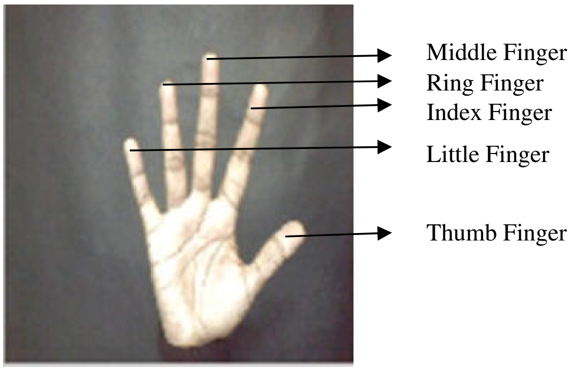


Figure 1. Name of fingers in hand

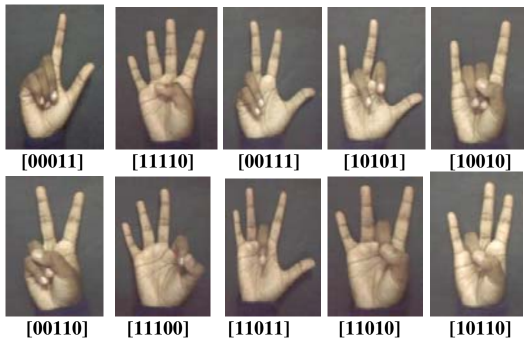
The tip of fingers is identified by measuring their heights with respect to a reference point at the bottom of the palm close to the wrist. The heights are determined by Euclidean distance measurements. Number of instances (denoted as 'd') in a scan are less than or equal to 3 in the case of Left-Right Scan and it is less than or equal to 2 in the case of Right-Left Scan which are determined by the 'UP' or 'DOWN' positions of the fingers. The hand pattern recognition procedure includes four major features. They are a) Data Acquisition b) Palm image Extraction c) Sign Detection d) Training Phase and e) Binary Conversion to Text as shown in Fig. 3.

A. Data Acquisition:

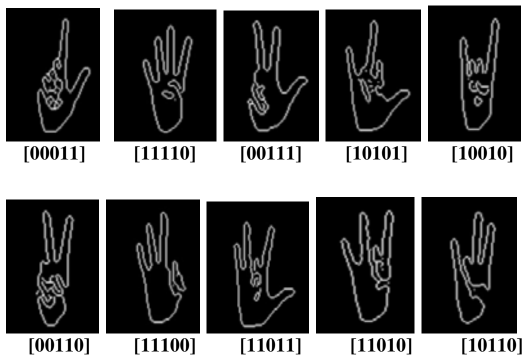
For the purpose of training the palm images, a maximum of 320 images, 10 each of the previously defined 32 signs are used. For the purpose of testing, 5 images of each of the predefined 32 signs are loaded one at a time, thus amounting to a total of 160 images. The images are captured at a resolution of 640 x 480 pixels. The runtime images for test phase are captured using USB web camera of LG Smart Cam. The images are captured in a high intensity environment directed to illuminate the image source which is held at black background so as to avoid shadow effects. The images are captured at a specified distance (typically 1.5 – 2 ft) between camera and signer. The distance is adjusted by the signer to get the required image clarity. In order to avoid the images of arm extension from palm, the images are taken with a black wrist band on the signers arm.

B. Palm Image Extraction:

The captured images are resized to a resolution of 128 x 128 pixels as shown in Fig 2 (a). The images in RGB colour are converted into gray scale images which are in turn converted into black and white image. The images are then processed using Canny Edge Detection technique to extract outline images (edge) of palm as shown in Fig 2 (b). It is easier to use those edge images for extracting the finger tip position for further processing.



(a)



(b)

Figure 2. Sample images of 32 signs
(a) Original images (b) Edge images

C. Sign Detection:

The edge images are further taken through scan process and detection phase. The procedure of scan process includes i) marking of feature points ii) determination of heights of fingers in 'UP' position iii) determination of angle between the line joining the feature point of 'UP' fingers with the reference point and the horizontal line passing through the reference point and iv) determination number of instances 'd' which is used to limit number of searches among the 32 signs is considered. In the testing phase, the details as computed above are then compared with an available static images data which were obtained through a training phase with a set of 320 images.

1) *Feature Point Extraction and Testing*: The edge images are scanned to locate the tip positions of the 'UP' finger and to locate the reference point (x_0, y_0) at the bottom of the palm. It requires either one of the two or both the scan modes namely the Left – Right and Right – Left scan to determine the feature

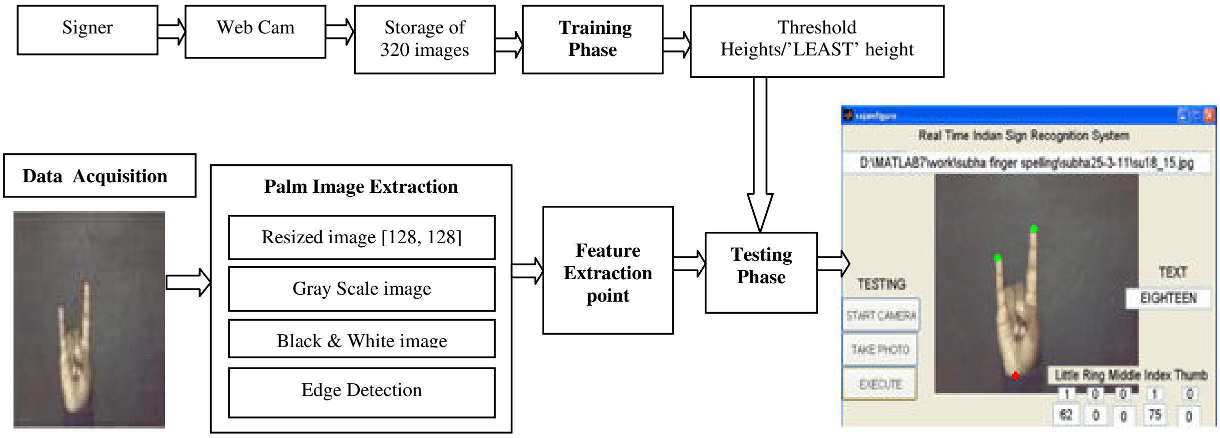


Figure 3. Proposed Methodology

points (less than or equal to 5 of fingers and 1 of reference point) as required by the sign under process out of the 32 signs defined. The feature points located by Left-Right scan are marked 'GREEN' color, those located by Right-Left scan are marked 'BLUE' color and the reference point is marked 'RED" color as shown in Fig. 4. The points extracted from edge images which are represented as (x, y) co-ordinates are marked with different colors in the resize image so that they become visible.

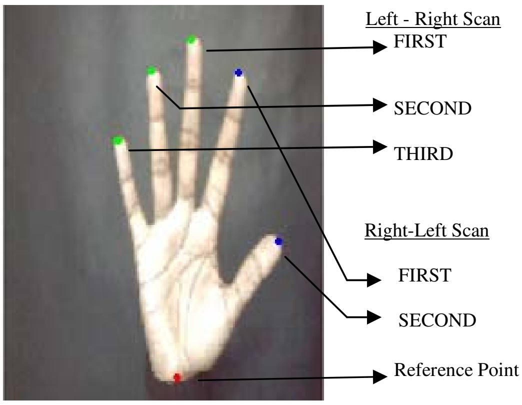


Figure 4. Color marked as red, Green, and blue in resized image

The origin point (0, 0) of the scan is the pixel at the top most left corner of the image space. The feature points are expressed with respect to this origin point as reference. The top most feature point located by Left-Right scan may be any of the five fingers. Locating of this 'FIRST' point will initialize a counter 'd' to 1. The following Left-Right scan will locate only those points to the left of the 'FIRST' point namely 'SECOND' and 'THIRD'. Each time a point is located; height is measured and compared with a preset height (say 'LEAST') obtained through training. If the measured height is greater than the 'LEAST' value, the counter 'd' is incremented by 1 up to a maximum of three. The maximum 'LEAST' height is obtained as the 'FIRST' point located with all the five fingers in 'DOWN' position. The 'UP' position of any finger is determined with reference to the 'LEAST height. Number of comparison to be made for identifying the fingers are fixed by the value updated in the counter 'd'. Analysis in the testing phase either uses the 'FIRST' point only or both 'FIRST' and 'SECOND' points depending on 'd'. The heights measured at these points are compared with threshold heights and thereby identify few or all of the five fingers.

The remaining fingers if any left unidentified in Left-Right scan necessititates further search into Right-Left scan which otherwise is set to immediately end up in the last pixel and thereby complete Right-Left scan.

Once a feature point is located, the scan process is directed to the next line (x_i to x_{i+1}). The new scan line is set to terminate with a margin of about 6 pixels from the y co-ordinate of the point located (i.e. y_i to y_{i-6}).

The origin of Right-Left scan is fixed with to the 'FIRST' located by Left-Right scan i.e. it starts from the line next to that of the 'FIRST' point. However, the margin along y co-ordinate direction is set from (y_i to y_{i+k}) where 'k' varies with the finger identified as the 'FIRST' point. This is done to avoid the search through the 'DOWN' fingers. Further scanning process and testing follow the same procedure as that of the Left-Right scan with the scanning process mode economical by setting the necessary margins and using the threshold as height required by the sign under study. The standard heights are the static data which are determined through the training phase. The scanning process stops when any of the following conditions is met.

- When 'd' = 3.
- When the point located is equal or less than preset 'LEAST' height which is obtained in training phase.

Testing: Comparison of the heights based on 'd' are fixed as shown in Table 1. Where $M1_L$, $I1_L$, $M2_L$, $I2_L$, $R2_L$, $L2_L$, $T2_L$,

$R3_L$, $L3_L$ are the threshold heights (minimum) of the five fingers obtained through training phase. There are slight differences in the heights of the same fingers for different signs. Therefore, the threshold heights are given as range between a maximum and a minimum instead of as a single height. Further, it is learnt from the study of images that based on 'd', these threshold heights of fingers except the 'THUMB' finger may be defined with 2 different ranges. For example $M1_L$, 'M' stands for 'MIDDLE' finger.

TABLE 1. COMPARISION OF THE HEIGHT BASED ON 'D' ARE FIXED

'FIRST' Point Located	**Order of Comparison**	**Remarks**
For d = 0	'LEAST'	When the FIRST point is equal or less than the preset 'LEAST' height.
For d = 3	$M1_L$, $I1_L$	The sequence of comparison identifies one of the two fingers namely 'MIDDLE' or 'INDEX'. This helps to directly fix the other two 'UP' fingers
For d = 2	$M2_L$, $I2_L$, $R2_L$	The sequence of comparison identifies one of the three fingers. To identify the next finger, the 'SECOND' point is used.
For d = 1	$M2_L$, $I2_L$, $R2_L$, $L2_L$, $T2_L$	The sequence of comparison identifies one of the five fingers.
'SECOND' Point Located	**Order of Comparison**	**Remarks**
For d = 2	$R3_L$, $L3_L$	The sequence of comparison identifies the second fingers which may be one of the two fingers.

'1' stands for the case when d = 3 (signs for which 'd' assumed a value between 0 and 3, the minimum value standard height has a different value represented by $M2_L$). Subscript '$_L$' stands for minimum (lower) value of the range defining the standard height of middle finger (In $M1_U$ represents the maximum height (upper) value of middle finger when d = 3).

Upper heights are used when there is an overlapping of ranges in some of the signs for e.g. the ranges for I2 and R2 were fixed as (71-77) and (67-72) respectively, representing ($I2_L$ – $I2_U$) and ($R2_L$ – $R2_U$). There is an overlapping of the heights between 71 and 72. This may lead to a wrong identification of fingers for some of the signs. Such an overlapping between fingers pairs whose heights are in the range such as 'MIDDLE' & 'INDEX', 'INDEX' & 'RING', AND 'RING' & 'LITTLE'. The upper limits are used to determine the heights that overlap between the ranges defined for such fingers.

For those signs where there is an overlap that leads to wrong identification, an additional measurement i.e. angle measurement helps to identify the correct finger of the finger pair listed above.

2) *Measurement of Heights:* Whenever a feature point say (x_1, y_1) located ('UP' position only), the corresponding heights with respect to reference point (x_0, y_0) are computed using Euclidean distance as in

$$\text{Euclidean distance} = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \quad (1)$$

where (x_1, y_1) is finger tip position

(x_0, y_0) is wrist hand position

(3) Measurement of Angle: The reference point (x_0, y_0) is in almost all the cases fixed at the centre of the bottom most scan line of the palm irrespective of the signer. The angle between the line joining finger tips with reference point and the horizontal line passing through the reference point differs with each of the five fingers. Fingers to the left of the 'MIDDLE' finger make obtuse angles and those fingers to the right of 'MIDDLE' finger makes acute angles. The palm is either held with fingers close to one another, or is stretched with maximum possible clearance between the neighbouring fingers. Accordingly the margins each finger varies with degree of stretch of palm image loaded. Because of the variation, angular measurement above is not to be used for recognition process.

D. Training Phase (Left – Right Scan):

Training phase determines the threshold heights of fingers using the 320 images corresponding to 32 signs of 10 images each. The 10 images of each sign are obtained with the palm held at a fixed distance from the camera between the fingers stretched at varying distances from each other. For each finger a minimum and a maximum threshold height are determined. The images of 32 signs are broadly divided into four categories, each of which consists of subset of signs.

Initially, the 'LEAST' height which serves as the reference threshold or 'UP' position is determined as given in section C.1

First category consists of LITTLE, RING in the 'UP' position and any one of the MIDDLE or INDEX fingers in the 'UP' position.

- With the MIDDLE finger in the 'UP' position, the remaining two fingers left unmarked amounts to 4 combinations i.e., 4 signs. 40 images corresponding to these 4 fingers are considered for determining the minimum and maximum threshold height ($M1_L$ &$M1_U$) of MIDDLE fingers.
- Likewise, with the INDEX finger in the 'UP' position, there is one finger left unmarked which makes up 2 combinations i.e. 2 signs; using these 20 images $I1_L$ & $I1_U$ are determined.

These 6 signs are not considered in the following process of threshold height determination. These heights are used for completing the Left-Right Scan for the case with d=3 (refer section C.1) during the testing phase.

In the second category threshold heights of all the five fingers ($M2_L$,$M2_U$,$I2_L$, $I2_U$,$R2_L$ $R2_U$, $L2_L$,$L2_U$, $T2_L$,$T2_U$) are determined. Each of these heights is determined using a subset of signs as detailed below

- With the MIDDLE finger ($M2_L$ & $M2_U$) in the 'UP' position combinations with the remaining 4 fingers amounts to 16 signs. Exempting the 4 signs considered in the first category, a group consisting of the remaining 12 signs (120 images) are considered.
- With the MIDDLE finger in the 'DOWN' position and INDEX ($I2_L$ & $I2_U$) finger in the 'UP' position, the remaining 3 fingers amounts to 8 signs (2^3). Exempting the 2 signs considered in the first category, a group consisting of the remaining 6 signs (60 images) are considered.
- With MIDDLE & INDEX in the 'DOWN' position and the RING finger ($R2_L$ & $R2_U$) in the 'UP' position, the remaining 2 fingers amounts to 4 signs for consideration.
- With RING, MIDDLE & INDEX in the 'DOWN' position and LITTLE finger ($L2_L$ & $L2_U$) in the 'UP' position, the only one THUMB makes 2 signs for consideration.
- With LITTLE, RING, MIDDLE, & INDEX in the 'DOWN' position and the THUMB finger ($T2_L$ & $T2_U$) in the 'UP' position, the number of consideration is just one.

With $1 < d < 3$, the 'FIRST' points located during testing in Left – Right Scan are compared with these second category threshold heights. Likewise the SECOND points located are tested with the third category heights that are determined as given below.

- The threshold height $R3_L$ & $R3_U$ are determined by considering the following 2 groups of signs.
  1. With LITTLE in the 'DOWN' and RING & MIDDLE in the 'UP' position, the remaining 2 finger amounts to 2^2 i.e. 4 signs for consideration.
  2. With LITTLE & MIDDLE in the 'DOWN' position and RING & INDEX in the 'UP' position, the only 'THUMB' finger that remain makes 2 signs for consideration.
- The threshold height $L3_L$ & $L3_U$ are determined by considering the following 3 groups of signs.
  1. With LITTLE & MIDDLE in the 'UP' position and RING in 'DOWN' position, the remaining 2 fingers amounts to 2^2 i.e. 4 signs for consideration.
  2. With LITTLE & INDEX in the 'UP' position and RING & MIDDLE in the 'DOWN' position, the only 'THUMB' finger that remains makes 2 signs for consideration.
  3. With LITTLE & RING in the 'UP' position MIDDLE & INDEX in the 'DOWN' position , the only 'THUMB' finger that remain makes 2 signs for consideration.

With fourth category of signs are grouped in such a way to determine the threshold heights $I4_L$, $I4_U$, $T4_L$, and $T4_U$ of INDEX and THUMB fingers using Right–Left Scan following the similar notions detailed above.

E. Binary Conversion to Text:

The output is obtained originally in the form of binary string of length of five in which the most significant bit represents the 'LITTLE' finger and the least significant bit represents the 'THUMB' finger. The string is then coded into the equivalent decimal numbers as shown in Fig 5. Each decimal number thus corresponds to a sign and may be assigned an equivalent Tamil letter.

IV. EXPERIMENTAL RESULTS

The proposed procedure was implemented and tested with 2 sets of images as given below

1. The same set of 320 images used for training using without angular measurements.
2. The new set of 160 images (5 images of each sign) loaded at run time using with angular measurement.

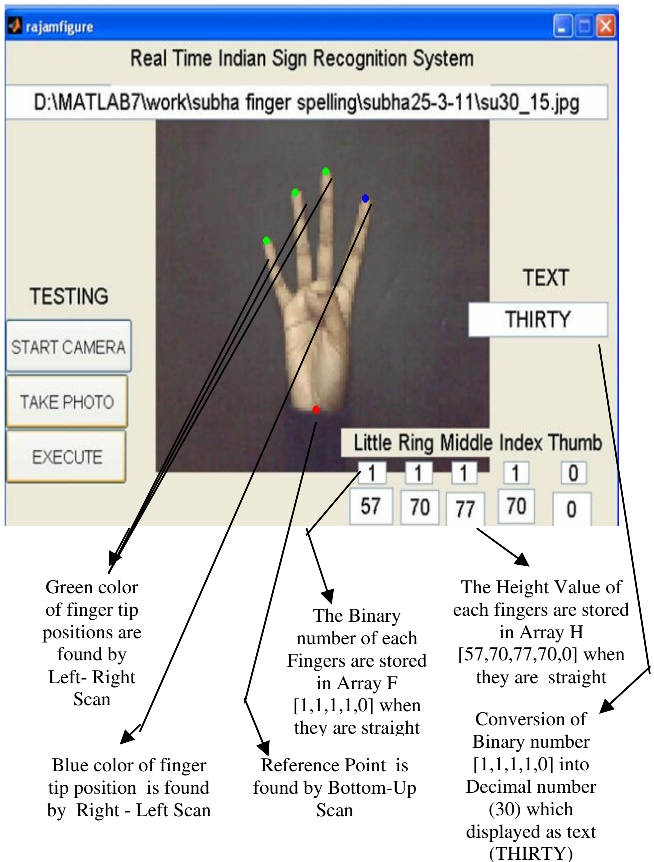


Figure 5. Sample Result of a Sign

Test with the first set of images produces exactly 100% recognition with almost all the 32 signs except 4 signs shown in Fig 6. For which the recognition percentage was about 70 to 80. Tests with the second set of images shows a significant improvement in recognition of signs. In this case, about 60 to 80 percentage recognition is obtained for 2 signs as shown in Fig 7. While producing 100% recognition with all other signs. The overall accuracy of the proposed method as given below

$$\text{Accuracy} = \frac{\text{No. of Patterns} - \text{No. of false result Patterns}}{\text{No. of Patterns}} \times 100\%$$

is obtained for both the test cases, Accuracy with first case test was 96.87% and that obtained with the second case was 98.125%.

Figure 6. Sample of false result patterns as first set of images

Figure 7. Sample of false result patterns as second set of images

V. Conclusions

A sign language recognition system proposed for human–computer interaction using Image Processing Technique was implemented successfully with accuracy comparable with those of recent contributions. The results presented shows that those signs producing lesser percentage of recognition can be improved by using the proposed angular measurement. The developed 32 signs can well be used to recognize one of the south Indian Languages i.e. the letters of Tamil language which has 12 vowels and 18 consonants. This paves scope for future work.

Acknowledgment

We would particularly like to thank Dr. A. Josphine Amala for stimulating discussion that we had. And also we thank Dr. B. E. Elizabeth Caroline for her valuable suggestions.

References

[1] 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.*

[2] 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, June 1999, Pages 291-309.*

[3] T. Shanableh, K. Assaleh, "Arabic sign language recognition in userindependent mode", IEEE International Conference on Intelligent and Advanced Systems 2007.

[4] T. Starner, J. Weaver, and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol.20, no. 12, pp. 1371-1375, Dec. 1998.

[5] M.W. Kadous, "Machine recognition of Australian signs using power gloves: Toward large-lexicon recognition of sign language," *Proc. Workshop Integration Gesture Language Speech*, pp. 165–174, 1996.

[6] J. S. Kim,W. Jang, and Z. Bien, "A dynamic gesture recognition system for the Korean sign language (KSL*)," IEEE Trans. Syst., Man, Cybern. B*, vol. 26, pp. 354–359, Apr. 1996.

[7] H. Matsuo, S. Igi, S. Lu, Y. Nagashima, Y. Takata, and T. Teshima, "The recognition algorithm with noncontact for Japanese sign language using morphological analysis," *Proc. Int. Gesture Workshop*, 1997, pp. 273–284.

[8] C. Wang, W. Gao, and Z. Xuan, "A Real-Time Large Vocabulary Continuous Recognition System for Chinese Sign Language," *Proc. IEEE Pacific Rim Conf. Multimedia*, pp. 150-157, 2001.

[9] K Assaleh, M Al-Rousan, "Recognition of Arabic Sign Language Alphabet Using Polynomial Classifiers," *EURASIPJournal on Applied Signal Processing*, vol. 2005, no. 13, pp. 2136-2145, 2005.

[10] T. Shanableh, K. Assaleh and M. Al-Rousan, "Spatio-Temporal feature extraction techniques for isolated gesture recognition in Arabic sign language *," IEEE Trans. Syst., Man, Cybern. B*, 37(3), June 2007.

[11] P. Subha Rajam, Dr. G. Balakrishnan, "Indian Sign Language Recognition System to aid Deaf – Dumb People",*IEEE International Conference on Computing Communication and Networking Technologies (ICCCNT), 2010, Pages(s):1-9.*

[12] 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.*

[13] W. Gao, G. Fang, D. Zhao, and Y. Chen, Transition Movement Models for Large Vocabulary Continuous Sign Language Recognition," *Proc. Sixth IEEE Int'l Conf. Automatic Face and Gesture Recognition, pp. 553-558, May 2004.*