Journal of Theoretical and Applied Information Technology
15th February 2012. Vol. 36 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

OFFLINE CANDIDATE HAND GESTURE SELECTION AND TRAJECTORY DETERMINATION FOR CONTINUOUS ETHIOPIAN SIGN LANGUAGE

ABADI TSEGAY[1], DR. KUMUDHA RAIMOND[2]

Addis Ababa University, Addis Ababa Institute of Technology
Addis Ababa, Ethiopia
E-mail: abadi_tsegay@hotmail.com[1] , kumudharaimond@yahoo.co.in[2]

ABSTRACT

A lot of effort has been invested in developing alphabet recognition and continuous sign language translation systems for many sign languages around the world. In this regard, little attention has been given to Ethiopian sign language (EthSL). However, an Ethiopian Manual Alphabet (EMA) recognition system has been developed in 2010. For a recognition system that can recognize continuous gestures from video which can be used as a translation system, a methodology that selects candidate gestures from sequence of video frames and determines hand movement trajectories is required. In this paper, a system that extracts candidate gestures for EMA and determines hand movement trajectories is proposed. The system has two separate parts namely Candidate Gesture Selection (CGS) and Hand Movement Trajectory Determination (HMTD). The CGS combines two metrics namely speed profile of continuous gestures for block division (BD) and Modified Hausdorff Distance (MHD) measure for gesture shape comparison and has an accuracy of 80.72%. The HMTD is done by considering the centroid of each hand gesture from frame to frame and using angle history, x-direction and y-direction of lines between successive centroids. A qualitative evaluation of the CGS is found to be 94.81%. The HMTD has an accuracy of 88.31%. The overall system performance is 71.88%.

Keywords: *EMA, Candidate Gesture Selection, Trajectory Determination, Modified Hausdorff Distance*

1. INTRODUCTION

Sign languages that exist around the world are usually identified by the country where they are used. For example, American Sign Language (ASL), Australian Sign Language (Auslan), British Sign Language (BSL), New Zealand Sign Language (NZSL), Taiwan Sign Language (TSL), Chinese Sign Language (CSL), Ethiopian Sign Language (EthSL) and so on.

Mostly, the communication among the hard of hearing people involves signs that stand for words by themselves. However, to make a sign language complete as a spoken language, the hard of hearing community around the world use manual alphabets for names, technical terms, and sometimes for emphasis. As there are different alphabets for different spoken languages such as English, Chinese, Greece, Ethiopia and so on, there are different types of manual alphabets or finger spellings used by the deaf people who use different sign languages.

In EthSL, there are 34 base alphabets called base or parent EMAs where each base alphabet has 6 other variations. In most sign languages other than EthSL, either there are alphabets that represent vowel sounds such as ***a, e, i, o*** and ***u*** of English language or there are limited back and forth movements to represent certain language accent as in Spanish Sign Language (SSL). However, in EthSL, the 6 variations are created or spelled by the same hand posture or form as the base alphabet and followed by unique hand movement trajectories for each variation. Figure 1 depicts one of the EMAs with the hand movement style and direction for the six variations of the alphabet.

Journal of Theoretical and Applied Information Technology
15th February 2012. Vol. 36 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

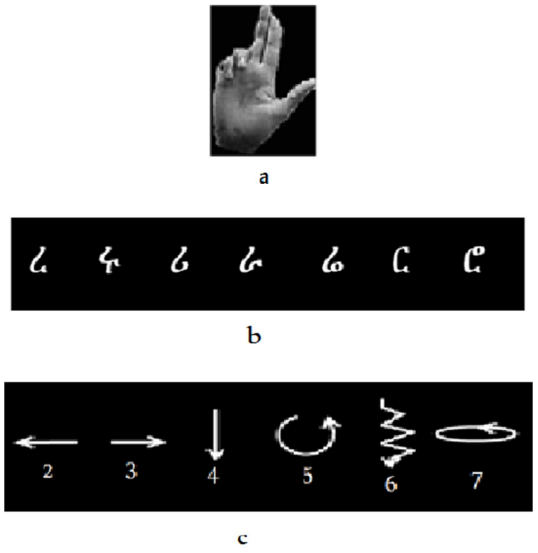

Figure 1. An EMA: a) Hand form for base EMA ( ) b) the 7 forms of ( ) c) Ideal trajectories for EMA

In Figure 1 (a), an EMA called (ذ) is shown while the 7 forms are depicted in (b). An ideal set of hand trajectories that are made when spelling the EMA forms other than the base EMA with a corresponding number is also shown in (c). In this report, the gestures are referred by the numbers shown below them in Figure 1 (c).

To analyze the trajectories for a specific object in video, there are three key steps:

- *Detection* of region of interest or object in each frame in the sequence.
- *Tracking* of such object from frame to frame using some property such as centroid of the object and generate a trajectory
- *Analysis* of the behavior of the trajectory

In [3], recognition of EMAs from static images has been developed using Gabor filter followed by PCA (Principal Component Analysis) and neural networks. Instead of recognizing EMAs only from static images, it is also a good idea to extract the EMAs and determine their trajectories from sequence of video frames. In this paper, candidate EMA frames are extracted from the video and the trajectories for the EMAs shown in Figure 1 (c) are determined. A recognition system may take the selected hand postures called candidates and trajectories as its input where a word level recognition is possible instead of an alphabet level.

An EMA recognition system has been developed in [3]. However, what if someone wants to interface the recognition system with a system that extracts sequence of candidate EMAs from video clips which can be used as input gestures to the recognition system? What if someone still wants to extend the ability of the recognition system to recognize not only the base EMAs but also the other variations of an EMA? If an EMA recognition system is interfaced with a system just described, it will be possible to translate a given video clip of a word "written" with EMA to voice. Therefore, the main purpose of this paper is to develop a technique which extracts important gestures called candidate gestures from a sequence of video frames and determines hand trajectories for each selected candidate gesture.

The rest of this paper is organized as follows. Some related works are described in section 2. Proposed design is discussed in section 3 and experimental results are presented in section 4. Conclusion is discussed in section 5 and future work in section 6.

2. RELATED WORKS

Extensive researches have been done on recognition of signs or finger spellings from static images. For example, Arabic Sign Language (ArSL) alphabet recognition system was built using polynomial classifiers as a classification engine for the recognition [1]. The researchers in [2] have also developed a system that recognizes alphabets in Persian Sign Language (PSL). EMA recognition system has also been developed in [3]. The research of sign language recognition is not limited to recognition from static images; it can also be applied to continuous pictures or videos of sign language [4, 5, 6, 7].

Sign language involves hand movements in different directions and hence translation systems for sign languages should be able to track the hand and determine the behavior of the trajectory of the hand movement. In [8, 9], hand gesture tracking system using Adaptive Kalman Filter was developed. The system in [8] segments the hand region using YCbCr color space and determines the hand position.

Usually, for CGS in a sequence of video frames, a key frame selection technique is applied. In [10], the researchers developed a novel method to extract key frames to represent video shot based on connectivity clustering. The method dynamically

Journal of Theoretical and Applied Information Technology
15th February 2012. Vol. 36 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

divides the frames into clusters depending on the content of shot, and then the frame closest to the cluster centroid is chosen as the key frame for the video shot. In [11], the researchers proposed an optimal key frame representation scheme based on global statistics for video shot retrieval. Each pixel in this optimal key frame is constructed by considering the probability of occurrence of those pixels at the corresponding pixel position among the frames in a video shot.

3. PROPOSED DESIGN

The first step in the proposed design is converting the given video into sequence of frames. A skin segmentation technique based on YCbCr color space is implemented with an explicit-definition skin-color modeling technique [13] to segment skin areas of the input images. The overall proposed design is shown in Figure 2.

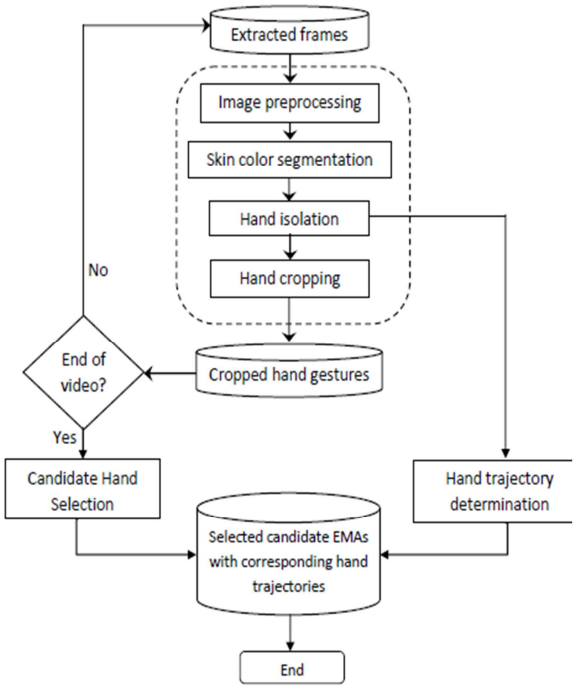

Figure 2. The Overall Proposed Design

Data were collected both for system design and system test. Five signers were considered and for the system to be general, different combinations of EMA variations were used. For each word either under the design or the testing column, two samples were recorded from each signer. Data collection is summarized as follows:

- For system designing
  - 7 words x 2 samples x 5 signers = 70 videos
- For system testing
  - 5 words x 2 samples x 5 signers = 50 videos
  - 6 words x 2 samples x 2 signers = 24 videos

Therefore, a total of 144 videos were collected to develop the proposed system where 70 of them were used to design the system and 74 of them were used to test the system. For the testing data, 2 samples for 5 words each were collected from 5 signers. Additional data of 2 samples for 6 words each were also collected from 2 signers.

Table 1. List of words used to design the system

Words used to design the system	
2 EMAs	***3 EMAs***
ሁዳ (HUDA)	ያሲን (YASIN)
ሳኒ(SUNNY)	ራህማ (RAHMA)
ባቲ (BATI)	
ዲላ (DILLA)	
ባሌ (BALLE)	

Table 2. List of words used to test the system

Words used to test the system	
2 EMAs	***3 EMAs***
ሙሉ (MULU)	ሐዋሳ (HAWASSA)
ኪያ (KIA)	ብሩክ (BIRUK)
ባላ (BALLA)	ሃይሉ (HAILU)
ካሴ (KASSIE)	ዳዊት (DAWIT)
ራያ (RAYA)	ኢዛና (EZANA)
ኩኩ (KUKU)	

The overall proposed design is composed of three main parts: ***skin-color segmentation***, ***candidate hand gesture selection*** and hand ***trajectory determination*** and each is discussed in subsequent sections.

Journal of Theoretical and Applied Information Technology
15th February 2012. Vol. 36 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1. Skin color segmentation

As object color segmentation is the most important stage in computer vision and sign language translation systems, in the proposed design, skin-color segmentation in the YCbCr color space with an explicit skin-color definition modeling [13] is implemented. Original images are shown in Figure 3.

Figure 3. Original RGB images: (a) and (b) from right-handed people, (c) from left-handed person

After applying the skin color segmentation to images (a), (b) and (c) in Figure 3, the proposed design results in the corresponding images of Figure 4.

Figure 4. Binary images after applying skin segmentation to images in Figure 3 (a), (b) and (c)

Images (a) and (b) in Figure 4 were taken under uncontrolled environment and that of (c) was taken under a controlled environment. After skin color segmentation, there are areas considered as skin while they are not. In fact, the areas detected falsely as skin are very small compared to the skin regions and these were removed by morphological operations. Opening followed by closing operations were used with a structuring element of d = 5 x 5 pixels.

3.2. Candidate Gesture selection (CGS)

3.2.1. Hand isolation

The objective here is to remove regions other than the hand region. This step assumes the hand region is always smaller than the face region. In the proposed design, very small regions were first removed using the morphological operations. Next, the area of each region was calculated in the segmented binary image and all areas were removed except ***the second largest***, which is usually the hand. With this unique approach, the proposed design is able to retain the hand either from the right- or left-handed person unlike the design in [11] which should be informed whether the signer is a right- or left-handed.

3.2.2. Block division

Before dividing the whole sequence of images into blocks based on speed profile of the gestures, the regions except the hand region were removed from all frames. And then the centroid of the hand region was collected from each frame in the sequence.

For the purpose of detecting significant motion of various hand movement speeds, centroids of alternating frames were considered for speed profile development. To find the speed of a hand between alternating centroids, the distance between the centroids is divided by twice the frame duration. Video camera used in this work recorded 30 frames per second (fps) and the frame duration is therefore 1/30 seconds (33.33 milliseconds).

Based on the speed profile developed for videos of the words listed in Table 1, a speed threshold of 0.155 pixels per millisecond was obtained for dividing the sequence of hand gestures into valid blocks. In Figure 5, a block division for the name SUNNY (ሳኒ) is displayed based on the speed threshold.

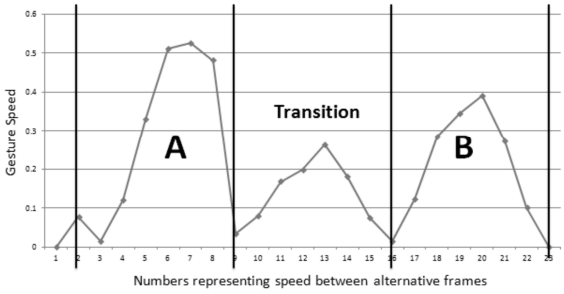

Figure 5. speed profile for a 2 EMA video clip for the name SUNNY (ሳኒ)

Blocks A and B are called EMA blocks and the central block is a transition between two valid EMAs.

3.2.3. Candidate hand gesture selection

After collecting centroid of each isolated hand gesture for hand tracking as described in section 3.2.2, the next task is to crop the hand gesture from the image frame. The proposed design searches for extreme white pixels and crop the hand based on

Journal of Theoretical and Applied Information Technology
15th February 2012. Vol. 36 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the extreme pixels either from right- or left-handed person as shown in Figure 6.

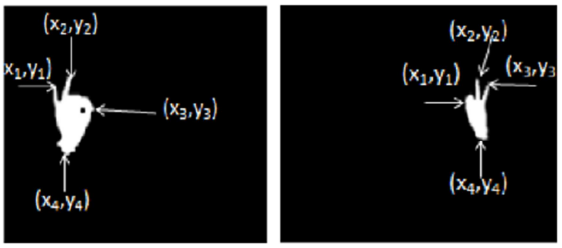

Figure 6. Binary images showing points of cropping for (a) Right-handed person and (b) Left-handed person

The cropped hand gestures (binary and corresponding gray scale gestures) are stored separately. The proposed design used the contour of the binary gestures for CGS and has selected the corresponding gray scale gesture.

To reduce the impact of signing errors, the search space was minimized by creating a search window within the EMA blocks as shown in Figure 7. In the x-axis of Figure 7, 3 represents the MHD between 3rd and 4th gestures, 4 represents the MHD between 4th and 5th gestures and so on. Search window for block A of Figure 5 is shown in figure 7. The proposed design localizes the search space for candidate selection in two steps:

- By dividing the whole sequence of image frames into blocks using the speed profile
- By creating a search window

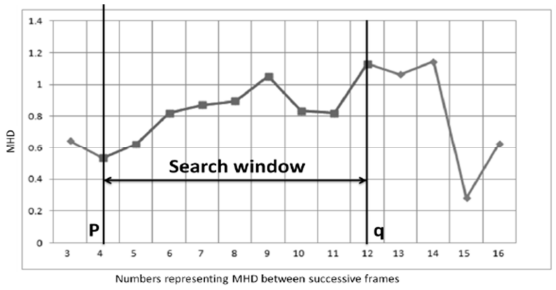

Figure 7. Search window definition within a block

The search window shown in Figure 7 starts at point p where $p = \frac{1}{10} * Block\ Size$ and ends at point q where $= \frac{2}{3} * Block\ Size$. These values were obtained experimentally such that the effect of transition gestures in the selection is minimized.

Within the created search window, a good candidate gesture is selected based on the minimum MHD between successive gestures. The first gesture that resulted in the minimum MHD is selected as candidate; if more than one value is obtained for the minimum variable, the first occurrence was used. Figure 8 shows the system output for candidate hand gesture selection of the name SUNNY (ﺱﻮﻧ). The selection was done in blocks A and B of Figure 5.

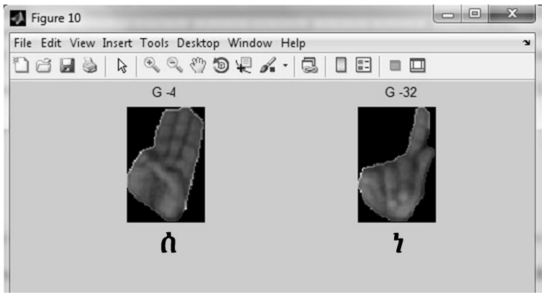

Figure 8. Output of the proposed system for candidate selection

Two candidate gestures were selected for base EMAs. The next section describes how to determine the behavior of the hand movement trajectory. When the behavior of the trajectory is determined, it is associated with the selected candidate gesture to come up with the actual EMA.

3.3. Hand trajectory determination

There are only 6 types of trajectories used in EMA where the 7th form is not considered in this work. As the types of the trajectories are very small, it is not feasible to use complex ways of hand tracking usually used in computer vision applications. Sample trajectories for EMAs are shown in Figure 9.

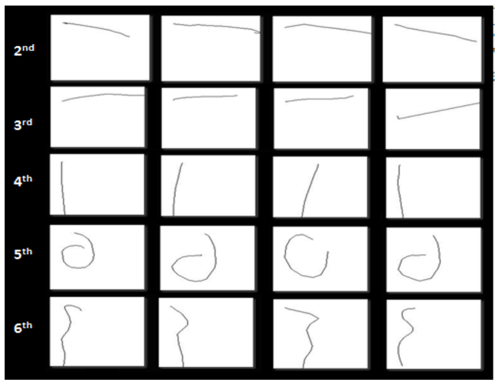

Figure 9. Sample trajectories

However, by carefully observing the nature of the trajectories used in EMA, in the proposed design, angle history and x- and y-directions of lines drawn between successive centroids were

Journal of Theoretical and Applied Information Technology
15th February 2012. Vol. 36 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

used as a cue to decide which trajectory is being formed using the signing hand. In general, the forms of the trajectories used in spelling EMAs are divided into four:

- Horizontal (straight) trajectory represents 2nd or 3rd forms
- Vertical (straight) trajectory representing 4th form
- Circular trajectory representing 5th form
- Vertical (zigzag or wavy) representing 6th form

The 5 types of the trajectories (2nd, 3rd, 4th, 5th and 6th) are discriminated from one another based on the following rules.

1. Calculate angles between successive centroids of trajectory T
2. If 75% of the absolute value of the angles is less than 45 degrees,
   2.1 Trajectory is straight horizontal
      2.1.1 Calculate *dd* which is the sum of all x-component distances between successive centroids
      2.1.2 If dd is greater than zero, T is 2nd or if dd is less than zero, T is 3rd.
3. Else if 75% the absolute value of the angles is greater than 60 degrees
   3.1 T is 4th.
4. Else if 75% of the absolute value of the angles is between 45 and 60 degrees
   4.1 Calculate yd's from y-component distance between centroids and calculate s, number of sign changes in x-direction between successive centroids.
      4.1.1 If 70% of yd's is greater than zero and s is greater than or equal 2, T is 6th
      4.1.2 Else if 30% of yd's is less than zero, T is 5th
5. End

4. EXPERIMENTAL RESULTS

4.1. Block division

Using the obtained speed threshold which is **0.155** pixels per millisecond, the proposed design was evaluated with 74 videos recorded for the words given in Table 2 for BD. Table 3 shows the experimental results of the proposed design for BD.

Table 3. Results of system test for BD

Signers	Number of videos	Threshold	Correct	Accuracy in %
1	10	0.155	9	90.00
2	10	0.155	9	90.00
3	10	0.155	8	80.00
4	22	0.155	18	81.82
5	22	0.155	19	86.36
Av.	**74**	**0.155**	**63**	**85.14**

4.2. Candidate Hand Gesture Selection

Five people were involved in the qualitative performance analysis of the proposed design. There were 154 EMAs and 770 observations from 5 people out of which 730 observations were correct. The experimental result for the CGS is presented in Table 4.

Table 4. Results of system test for CGS

Signer	Number of EMAs	Number of people	Total observation	Correct	Accuracy in %
1	23	5	115	110	**95.65**
2	23	5	115	95	**82.61**
3	20	5	100	90	**90.00**
4	42	5	210	210	**100.0**
5	46	5	230	225	**97.83**
Av.	**154**	**5**	**770**	**730**	**94.81**

As can be seen from Tables 3 and 4, the CGS module has an accuracy of 85.14% * 94.81% = 80.72%.

4.3. Hand Movement Trajectory Determination

Before cropping the hand gesture, the centroid of each gesture was collected for the purpose of determining hand trajectories. The experimental result for this module is shown in Table 5.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5. Experimental result of system test for HMTD

Signers	Number of EMA trajectories	Correct trajectories	Accuracy in %
1	23	21	**91.30**
2	23	17	**73.91**
3	20	16	**80.00**
4	42	39	**92.86**
5	46	43	**93.48**
Av.	**154**	**136**	**88.31**

4.4. Overall system performance

As shown in Table 4.3, out of 154 trajectories, 136 were found to be correct. However, the quality of the gesture whose trajectory is regarded as correct may not be found to be satisfactory. Therefore, multiplying the CGS by the HMTD is a wrong approach. So a separate analysis of the system output both for the qualitative (CGS) and quantitative (HMTD) gives a different result. First, for each correct output of the BD, output of CGS and that of HMTD should be correct altogether. And the percentage of the correct occurrences is multiplied with the BD. Hence, even if there are 146 qualitatively satisfactory gestures and 136 correct trajectories, only 130 of them are correct in both the qualitative result and the HMTD. So the overall system performance is:

Overall Performance = 130/154 * 85.14
= 84.42% * 85.14
= 71.88%

5. CONCLUSION

In this paper, a system that extracts important hand gestures that represent valid EMAs and determines their hand movement trajectories is developed. The system comprises image segmentation, blob analysis for hand isolation, CGS, and hand movement HMTD. YCbCr color space is used for skin-color segmentation because the discrimination between skin and non-skin pixels in the Cb-Cr plane is very good and the computational expense is less. With the color space used, the system is able to segment the image effectively. Following the segmentation is the blob analysis for hand isolation.

The hand isolation module heavily depends on the segmentation stage because it considers blob sizes to isolate the hand. If the segmentation is poor, the hand isolation will also be poor. With the data collected for system design and test, both the skin-color segmentation and the hand isolation are effective.

The CGS uses a speed profile of gestures to divide a sequence of image frames of a video into blocks that represent EMAs and return paths. This approach localizes the candidate search space to a smaller block than to the whole sequence of images. After dividing sequence of image frames into blocks, a search for EMA is done in alternating blocks. The proposed design introduces a search window for effective candidate search and has an accuracy of 94.81%. The overall accuracy of the CGS module is 80.72%.

The HMTD module uses the angle and direction information from the centroids with a given EMA block. For final decision of trajectory, angle history between successive centroids is computed and the x- and y-directions of the trajectories are considered. This module has an accuracy of 88.31%.

The proposed design works fine for all EMA forms except for 1st and 7th. The first is “spelled” just with a static EMA symbol without any hand movement. The 7th is also “spelled” with an EMA symbol where there is only a rotating hand movement with a fixed axis. In both cases, the proposed method, as it uses a hand movement speed as clue, is not able to extract those EMA forms.

6. FUTURE WORKS

In this paper, candidate hand gesture selection and trajectory determination for the selected EMA is done. The outputs of the proposed design can be used as an input to a recognition system where a word or sentences level recognition is required.

The signers have been oriented to avoid overlapping between hand and face. Therefore, to make the signers free from this constraint, the concept of digital image processing called occlusion can be used to separate overlapping objects or else colored gloves can be used for a good segmentation even when there is overlapping between hand and face.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES

[1] Khaled Assaleh and M. Al-Rousan, "Recognition of Arabic Sign Language Alphabet Using Polynomial Classifiers", EURASIP Journal on Applied Signal Processing 2005:13, 2136-2145, 2005

[2] Azadeh Kiani Sarkaleh, Fereshteh Poorahangaryan, Bahman Zanj and Ali Karami, "A Neural Network Based System for Persian Sign Language Recognition", IEEE International Conference on Signal and Image Processing Applications, 2009.

[3] Yonas Fantahun Admasu and Kumudha Raimond, "Ethiopian Sign Language Recognition Using Artificial Neural Network", 10th International Conference on Intelligent Systems Design and Applications, 2010.

[4] Yang quan, "Chinese Sign Language Recognition Based on Video Sequence Appearance Modeling", 5th IEEE Conference on Industrial Electronics and Applications, 2010

[5] Maryam Pahlevanzadeh, Mansour Vafadoost, Majid Shahnazi, "Sign Language Recognition", http://www.osun.org.

[6] Mohamed Mohandes and Mohamed Deriche, "Image based Arabic Sign Language recognition", 6th IEEE Trans. 978-1-4244-5046, 2010.

[7] Justus Piater, Thomas Hoyoux, Wei Du, "Video Analysis for Continuous Sign Language Recognition", 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, 2010.

[8] Nguyen Dang Binh, Enokida Shuichi and Toshiaki Ejima, "Real-Time Hand Tracking and Gesture Recognition System", GVIP 05 Conference, CICC, Cairo, Egypt, 19-21 December 2005.

[9] Mohd Shahrimie Mohd Asaari and Shahrel Azmin Suandi, "Hand Gesture Tracking System Using Adaptive Kalman Filter", IEEE 978-1-4244-8136, 2010.

[10] Yongliang Xiao and Limin Xia, "Key Frame Extraction Based on Connectivity Clustering", Second International Workshop on Education Technology and Computer Science, 2010.

[11] Kin-Wai Sze, Kin-Man Lam, and Guoping Qiu, "A New Key Frame Representation for Video Segment Retrieval", IEEE transactions on circuits and systems for video technology, vol. 15, No. 9, September, 2005.

[12] Sánchez-Nielsen, Luis Antón-Canalís and Mario Hernández-Tejera, "Hand Gesture Recognition for Human-Machine Interaction", Journal of WSCG, Vol.12, No.1-3, WSCG'2004, Plzen, Czech Republic, February 2-6, 2003.

[13] Simple Skin Segmentation, http://www.mathworks.com/matlabcentral/fileexchange/24934-simple-skin-segmentation/all_files

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AUTHOR PROFILES

Abadi Tsegay Weldegebriel received BSc. in Electrical Engineering from Mekelle University, Ethiopia, in 2006. He has also received MSc. in Computer Engineering from Addis Ababa University in 2011. Currently, he is a lecturer at Hawassa University, Ethiopia. His research interests are Image processing, Computer vision and Artificial intelligence.

Dr. Kumudha Raimond received her B.E from Madras University and M.E from Government College of Technology, Coimbatore and Doctoral degree from Indian Institute of Technology, Madras, India. She is having twelve years of teaching experience along with three years of industrial experience in General Electric, India. Her research interests are intelligent systems, adhoc protocols, wireless sensor networks, image processing, compression, watermarking and biometric, biomedical and bioinformatics applications.