ASL Finger Spelling Recognition System for Interactive Learning and Education Purpose

J. H. Koh and S. H. A. Ali
Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400 Batu Pahat, Johor, Malaysia.
aminahh@uthm.edu.my

***Abstract*—The learning platforms for sign language are very limited. Images and videos are usually used to deliver the process of sign language learning. However, sign language is complicated and hard to practice. Thus, a vision-based American Sign Language (ASL) finger spelling recognition system is developed and presented in this paper. The aim of this project is to help users to learn and practice the sign language. ASL is the most common used sign language around the world. Five ASL finger spellings are used as the data set in this project which are 'A', 'E', 'I', 'O' and 'U'. MATLAB software was used to design the graphical user interface (GUI) of the system. The GUI is divided into 3 parts which are Learn, Test and Quiz. The input image was captured by webcam and then converted into HSV color space for segmentation. Recognition process was done by computing 2D correlation coefficient between input testing images and database training images. The accuracy and processing time of 500 testing images with 100 training images are 97.8% and 0.9434 seconds respectively. The challenges of vision-based hand gesture recognition system such as the complexity background and background color were analyzed and they are presented in this paper.**

***Index Terms*—ASL Finger Spelling; Hand Gesture Recognition; MATLAB.**

I. Introduction

Hand gesture is the movement of hand to express ideas and meanings. Gesture recognition enables human to communicate and interact with the machines without any additional mechanical devices. For example, a user is able to use hand gesture to control the movement of mouse cursor [1] [2]. Besides, hand gesture recognition is able to be employed in other applications such as sign language, robot control, tracking gestures and gaming [3].

Learning experience is the key element to produce capable citizen for our country. Due to the peak development in science and technology fields, there are some electronic dictionaries have been created to assist people in learning new words in various languages. The most common and well-known translation platform is Google Translate. It detects the input words or sentences and translates them into words or sentences in the desired language. In learning the sign language, there is a system limitation exists in the process of learning as most of the sign language learning platforms only use pictures and videos to deliver information.

Great attention should be given to the sign language learning since it is an effective way to communicate among people with profound hearing and speech impairment. Everyone deserves a chance to learn in a good environment, so does the people with profound hearing and speech impairment. A vision-based hand gestures recognition system should be implemented in their learning process in order to enhance their understanding. Besides, the recognition system is able to help beginners to learn and practice the sign language in a more convenient way than the current one.

There are various types of sign language exist around the world. The sign language involves hand gestures, body movement and facial expressions to express the thoughts of people. It is primarily used by the profound hearing impairment community and others such as speech impaired community or people with difficulty in using spoken language due to some other disabilities. The sign language is not a universal language as different sign language is used in different country just like the spoken language all over the world [4]. American Sign Language (ASL) is the most widely used sign language around the world. Therefore, ASL finger-spelling was chosen as the data set in this project.

Basically, the system consists of three major parts which are (1) A webcam that captures the hand gestures images and saves it in the system, (2) MATLAB that sets up the fundamental function of the system, which are image segmentation, extraction and recognition and (3) lastly, an attractive graphical user interface (GUI) that visualizes the real-time hand gesture detection and the output of the recognition results.

Throughout the project, a GUI will be designed and developed. The GUI is divided into 3 parts which are Learn, Test and Quiz. Users learn the sign language by referring to the images provided followed by testing or practicing the sign language via the real-time hand gesture recognition. Quizzes are also provided to keep the learning interesting.

Although hand gesture technology is considered as a highly successful technology by making human life to be simple, there are still many challenges associated with the accuracy of the system can be found. Even though the vision-based hand gesture recognition is more user friendly and inexpensive than the hardware implementation hand recognition system, it has lower accuracy than the mechanical based system due to the existence of image noise problem. The presence of inconsistent lighting and background environment makes the recognition process to be difficult. Moreover, the quality and resolution of camera can also cause variation to occur in recognition accuracy [5].

II. System Design

Five American Sign Language (ASL) finger spellings as shown in Figure 1 are used as the training set of the project, which are 'A', 'E', 'I', 'O' and 'U'. A total of 100 training images were collected from 10 different persons with different size of palm, different length of fingers and different skin color. Each person contributed 2 training images to each finger spelling gesture. The purpose of having training images from 10 different persons is to increase the accuracy of the system when it is tested by different person. The training images were taken by using the same algorithm of the hand gesture testing system without going through the recognition process. All training sets were captured with plain white background and then they were converted into binary images. The training images were saved in the size of 256x256. White pixels represent the hand gesture whereas the black pixels represent the background. Figure 2 shows the examples of training images.

Figure 1: ASL finger spelling data set

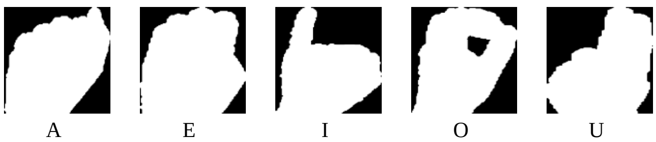

Figure 2: ASL Finger Spelling Training Images

III. SOFTWARE DEVELOPMENT

A webcam was connected to the MATLAB software. The video was set to manually trigger and capture 1 frame per trigger. This project was considered as a real time continuous system which captured the testing image in every 10 seconds and then processed it. Users located their testing hand gestures inside a designed boundary box in order to reduce the complexity of the background and simplify the image processing. Only the images within the boundary box were extracted and would be further processed. Figure 3 illustrates the testing situation. The complete flow chart of the main program is shown in Figure 4.

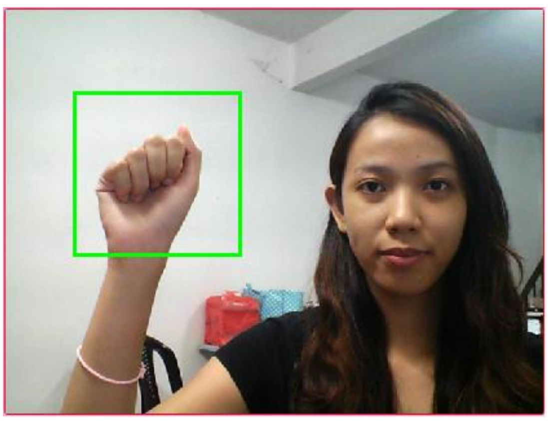

Figure 3: Testing Situation with Boundary Box

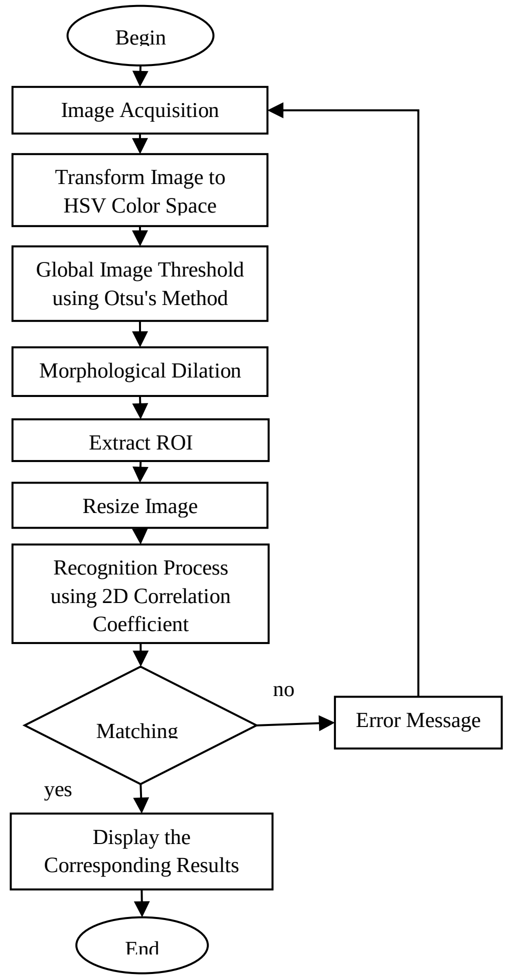

Figure 4: Flow Chart of The System

There are 3 main phases of a hand gesture recognition system which are segmentation, extraction and recognition. Global image threshold by using Otsu's method [6] was implemented on Saturation Channel in HSV color space to segment the region of interest (ROI) from the background. After filtering noise by morphological dilation, there might be some small contours present in the images in which they are not the ROI. To ignore the unwanted contour, an assumption that the hand contour is the largest contour was made. Finally, the extracted ROI image was then resized to 256x256 which is the same size as the training images to make the gesture recognition invariant to image scale. Figure 5 illustrates the output of each image processing stage.

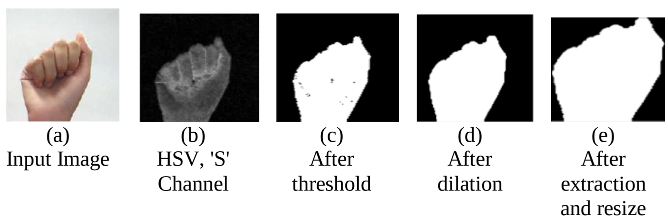

Figure 5: Output of Each Image Processing Stage

A 2D correlation coefficient was then used in matching and recognizing the sign language. The 2D correlation coefficient of the captured testing image and each of the database training image were computed. The correlation coefficient value lay between the range of 0 and 1. The training image with maximum value of 2D correlation coefficient with the testing image was chosen as the final recognition answer.

IV. Results and Analysis

This chapter explains about the parameters tuning, data collection method and the results of GUI. All testing results and processing time were recorded to calculate the accuracy and total average processing time.

A. *Background Color Analysis*

To choose the best background color, a few tests were carried out. A total of four color was chosen for the test which were red, green, blue and white. Red, green and blue are the primary colors which could be combined in various ways to produce other colors while white is made by mixing red, blue and green light at full intensity. The experiment testing was done by using color board as the background. Apart from that, the testing had also included different light intensity environments which were low light intensity and high light intensity. For the investigation using low light intensity (48 lx), the testing was carried out in an indoor room during day time without any additional light sources. For the study using high light intensity (152 lx), the testing was carried out in an indoor room during day time with an additional fluorescent tube lighting source turned on.

Before starting the experiment, the testing images of each channel in HSV color space were obtained and compared. The channel image with the highest contrast between the background and the ROI was then selected to undergo the process of dilation and extraction. Some of the channels were inverted to make the background to be in black pixels and the ROI to be in white pixels so that they were as similar as the database image. By using the best selected channel, the best edge detection and most accurate output are obtained. Table 1 states the best selected HSV channel for the segmentation by using different background color in two different light intensity levels whereas Table 2 explains the summary of the background color analysis in both light intensity conditions.

Table 1
The Best Selected HSV Channel for Segmentation

	Red	Green	Blue	White
Low Light Intensity	Inverted 'S' Channel	Inverted 'S' Channel	Inverted 'H' Channel	S Channel
High Light Intensity	Inverted 'S' Channel	Inverted 'S' Channel	Inverted 'S' Channel	S Channel

Based on the summary, the results are shown better in the higher light intensity level than the low one. Blue and white colors perform the best in both light level conditions. However, the selected channel for blue color background is different in these two conditions. It indicates that the testing results by using blue color background is not consistent. Different light intensity could affect the properties of the images when using the blue color as a background.

Theoretically, white light is made up of seven different colors namely, red, orange, yellow, green, blue, indigo and violet. The red, green and blue color boards will reflect and absorb some colors from the light causing an uneven color distribution on the board. For instance, if there is only blue light shone onto a red shirt, the blue light is absorbed and there is no red light to be reflected. Thus, the shirt appears as black [7]. Therefore, white color was chosen as the best testing background because white color reflects all the color and remains white.

Table 2
Summary of Background Color Analysis

	Red	Green	Blue	White
Low Light Intensity				
	Small holes are detected. Light reflection on color board is considered as part of ROI.	Perfect	Perfect	Small holes are detected.
High Light Intensity				
	Perfect	Large holes and some noises are detected.	Perfect	Perfect

B. *Experiment Testing*

Malaysia is a multi-racial country that consists of different ethnic groups. The three main ethnic groups in Malaysia are Malays, Chinese and Indians. Therefore, the testing was carried out by having 20 volunteers who are the students of Universiti Tun Hussein Onn Malaysia (UTHM) in a combination of 8 Malays, 8 Chinese and 4 Indians. A total of 100 testing images was taken for each finger spelling gestures. The skin colors of the three main races groups in Malaysia are slightly different. Typically, Malays have olive skin color, Chinese have fair skin color and Indians have dark skin color. Table 3 shows the overall accuracy results of five different ASL signs for Malays, Chinese and Indians.

Table 3
Accuracy Result from Different Skin Colors

	A		**E**		**I**		**O**		**U**	
	√	X	√	X	√	X	√	X	√	X
Malay	39	1	39	1	40	0	38	2	40	0
Chinese	39	1	40	0	40	0	37	3	40	0
Indian	20	0	18	2	20	0	19	1	20	0
Total	98	2	97	3	100	0	94	6	100	0
Accuracy (%)	98		97		100		94		100	

From Table 2, there are a total of 11 images was misrecognized out of 500 testing images. The accuracy percentage for ASL finger spelling 'A', 'E', 'I', 'O' and 'U' are 98%, 97%, 100%, 94% and 100% respectively. The overall accuracy percentage is 97.8%. The skin color of different ethnic group does not have any significant effect on the accuracy as the testing was carried out in white color background. The contrast is still high between the white color background and different skin color. Thus, the background and ROI could be properly segmented.

For the misclassified finger spelling, there are 2 finger spellings of 'A' are misclassified as 'E' and 3 finger spellings of 'E' are misclassified as 'A'. Besides, there are a total of 6 misclassified finger spellings of 'O' with 3 of them are misclassified as 'A' and another 3 are misclassified as 'E'. The reason of the misrecognized finger spelling is due to the captured angle. The hand gesture must always be vertical and directly faced towards the webcam. Once the orientation angle is slightly different, the outline of the hand gesture will change and affect the results leading to misclassification.

C. Graphical User Interface

Figure 6 shows the GUI of the home page. The center of the home page shows the ASL finger spelling image. There are 3 push buttons which are 'LEARN ', 'TEST' and 'QUIZ'. Each push button redirects the system to the corresponding page.

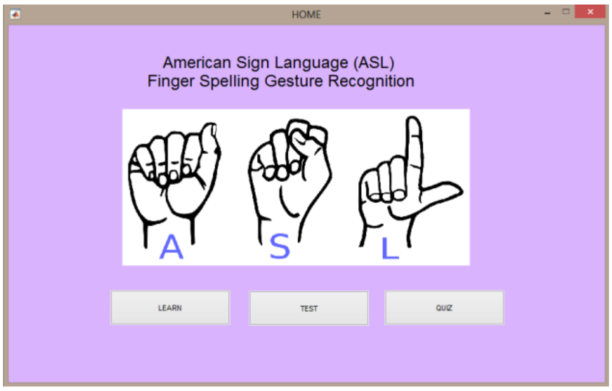

Figure 6: GUI of Home Page

Figure 7 shows the GUI of the LEARN page with finger spelling of 'A'. The purpose of this page is to allow the user to learn the ASL Finger Spelling. There are 5 available finger spellings in the drop down list which are 'A', 'E', 'I', 'O' and 'U'. Once the users click on an alphabet in the drop down list, the corresponding finger spelling image will be displayed on the right hand side.

Figure 8 shows the GUI of the TEST page. This page is mainly focused on the real time ASL finger spelling testing purpose. Users must click on the 'Start' button to begin the testing and 'Stop' button before quitting this page. The system captures the testing image in every 10 seconds as displayed in Time (sec) in the GUI. The period of 10 seconds is provided to allow the users to be ready and to locate their hand gestures inside the designed boundary box. The real time preview is shown in Axes 1 and the captured testing image is shown in Axes 2. Finally, the results of the recognized finger spelling image and the alphabet are displayed in Axes 3 and Edit box 1 respectively. Since the system is a continuous system, Edit box 2 is added to display cumulative alphabets results.

Figure 9 shows the GUI after completing one test. The processing time between the testing images is captured and the time of the displayed recognition results is measured and shown in the GUI.

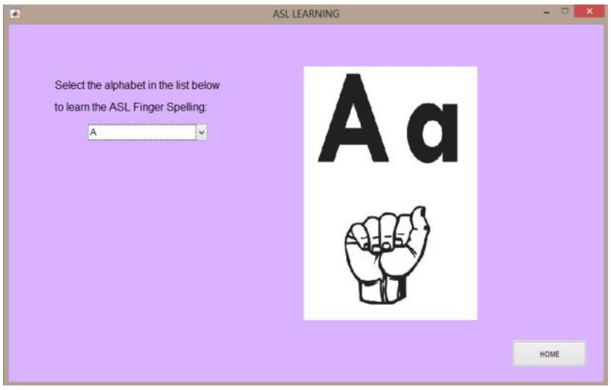

Figure 7: GUI of Learn Page

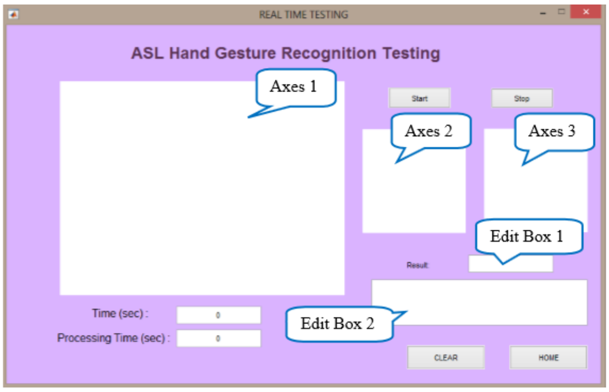

Figure 8: GUI of Test Page

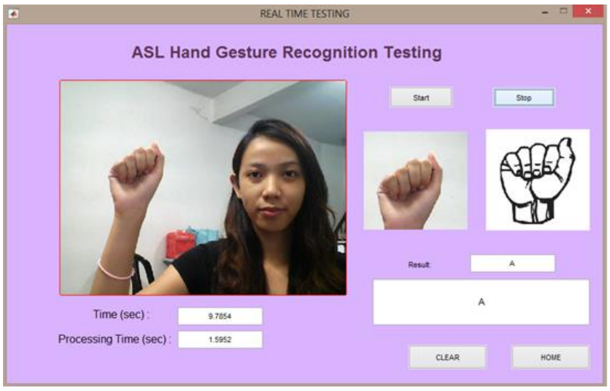

Figure 9: Results on Test Page

Quizzes are designed to keep the learning interesting. Figure 10 shows the GUI example of the quiz page. Users have to answer the questions by filling in the correct missing alphabets. In order to fill in the answer, they must click on the missing alphabet. Once they click on the missing alphabet, the ASL finger spelling recognition interface will be shown and started as illustrated in Figure 11. Similar to the testing system as described on Test page, the recognition system will capture the input image once it reaches 10 seconds.

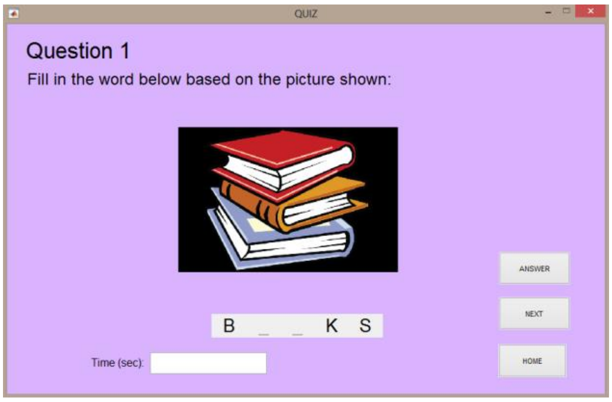

Figure 10: GUI of Quiz Page

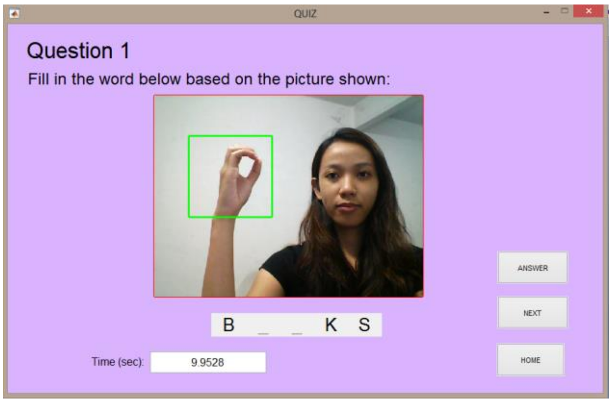

Figure 11: Answering Quiz

V. CONCLUSION

The vision-based ASL finger spelling recognition system is able to help users to learn and practice sign language in a more convenient way. The system captures the input image from a webcam, processes it, matches it with the database, and finally classifies it to its belonging sign language. This system is not only beneficial for the group with profound hearing and speech impairment, but also for people who are interested in learning the sign language. Besides, this system has the potential to reduce the social gap between the profound hearing and speech impaired group and other society thus achieving harmony equality.

The performance of the ASL finger spelling recognition system was evaluated on a data set of 500 testing images. The experimental results show that this approach performs well and it is fit for the real time application. The obtained accuracy is 97.8% and the average processing time is 0.9434 seconds.

However, the performance of the system highly depends on the background environment and database training images. If there are other objects exist in the testing background, it will affect the result of the hand gesture recognition thus degrading the performance. Besides, different background color requires different algorithms to perform the hand gesture recognition. The database training images have to be taken from various positions and angles in order to increase the accuracy.

The usage of camera instead of normal webcam is highly recommended so that a high-resolution image can be generated thus simplifying the image processing. For future recommendation, the hand features such as palm, wrist line, fingers can be extracted to increase the robustness of the system by having rotation calibration.

ACKNOWLEDGMENT

The authors would like to thank University Tun Hussein Onn Malaysia (UTHM) for providing the facilities and fund to complete this project.

REFERENCES

[1] S. Barve, K. Dhatavkar, S. Singh, V. Baghele and A. Jethwani , "Hand Gesture Recognition Method for Human," *International Journal of Information & Computation Technology*, vol. 2, no. 2, 2012, pp. 57-62.

[2] S. Salian, P. Ganorkar and D. Serai, "Hand Gesture Recognition and Cursor Control," *ResearchGate*, Mumbai, 2015.

[3] R.Z. Khan and N.A. Ibraheem, "Survey on Gesture Recognition for Hand Image Postures," *Computer and Information Science*, vol. 5, no. 3, 2012, pp. 110-121.

[4] Z. Jachova, O. Kovacheva and A. Karovska, "Differences Between American Sign Language (ASL) and British Sign Language (BSL)," *Special Education and Rehabilitation*, vol. 9, 2008, pp. 41-54.

[5] Ankita P. Meshram , Dr.D. V. Rojatkar, "Gesture recognition technology," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 4, no. 1, 2017, pp. 135-138.

[6] N. OTSU, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, 1979, pp. 62-66.

[7] "Colours of light," Science Learning Hub, [Online]. Available: https://www.sciencelearn.org.nz/resources/47-colours-of-light. [Accessed 30 4 2017].