Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

MALAYSIAN SIGN LANGUAGE DATASET FOR AUTOMATIC SIGN LANGUAGE RECOGNITION SYSTEM

M. Karbasi, A. Zabidi, I. M. Yassin*, A. Waqas and Z. Bhatti

Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

Published online: 05 October 2017

ABSTRACT

Hearing impaired individuals have issues to communicate with normal people. They have their own language called Sign Language (SL) to express their feeling or to communicate with others. As communication is an essential part of normal everyday life, it is particularly important for deaf people to communicate as normally as possible with others. Recent advancements in computing technologies have the potential to be applied in the field of SL recognition. These computer-based approaches are able to translate the SL into verbal language and vice-versa. This paper describes the development of a dataset for an automated SL recognition system based on the Malaysian Sign Language (MSL). Implementation results are described.

Keywords: sign language; pattern classification; database.

Author Correspondence, e-mail: ihsan.yassin@gmail.com

doi: http://dx.doi.org/10.4314/jfas.v9i4s.26

1. INTRODUCTION

Hearing impaired individuals have issues with communication with normal people. They have their own language called Sign Language (SL) to express their feeling or to communicate

Journal of Fundamental and Applied Sciences is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. Libraries Resource Directory. We are listed under Research Associations category.

with others. This form of communication involves hand movement and additional facial expressions as well. In Malaysia, the number of people who understand SL is small with approximately 40,000 deaf population registered with the Social Welfare Department of Malaysia by late December 2011 [7].

Malaysia had signed the United Nations Convention on the Right of People with Disabilities (UNCRPD) and passed the People with Disability Act 2008 (Act 685) to give the people with disabilities the opportunity to live as normal citizen of Malaysia [4]. As communication is an essential part of normal everyday life, it is particularly important for deaf people to communicate as normally as possible with others. However, there is a lack of qualified SL interpreters (SLI) in Malaysia and high demand for their services. This problem is further aggravated by the small number of normal people that understand this form of communication, making SL communication very difficult [8].

SL has a complex structure with consists of a combination of manual gestures (movement and orientation of hands and arms [23]) and non-manual gestures (mainly facial expression, head movement, posture and orientation of body and torso) that convey lexical meaning [11]. A type of SL called the Malaysia Sign Language (MSL), a derivation of the American Sign Language (ASL) is commonly used for communication in Malaysia [3].

Progration in innovation technology (for example, video processing, AI, Human Computer Interaction (HCI) system) could be possibly connected to SL. Computer programming could help to reduce whose are not expert in SLI by applied intermediate system between deaf and normal people. The target of such system is to translate signs to text and voice to text as well. The aim of such applications are to help deaf people study in normal school. Such these systems are active research area which implemented by [3, 12-13, 19].

Microsoft Kinect is a sensor developed by Microsoft Corporation for use with its Xbox 360 gaming platform. The device features an Infrared (IR) depth transceiver that is able to approximate depth information from subjects in front of the camera and its ability to generate a three-dimensional depth map of its surroundings. Due to this interesting capability, many researchers have utilized the Kinect sensor to perform acquisition and image processing-based tasks. Several works in this area are in [1-2, 10] has been done. Therefore, our research goal is to utilize this sensor to perform research on its potential application in SL recognition.

The first step to train any intelligent system is to collect the data necessary in order for the system to learn specific characteristics in the data, which can be used to infer and construct knowledge. Therefore, in this paper, we present the development of a MSL dataset consisting of static and dynamic signs for training of an artificial [17] intelligence-based system for SL recognition. The static signs consist of static images depicting hand gestures of alphabets, while the dynamic signs consist of SL gestures indicating certain words and phrases in the English language. In our work, we collect not only Red Green Blue (RGB) images but also depth and skeletal information to track the SL shapes/textures and gestures. We obtained from the Kinect that has the ability to extract the aforementioned features using its built-in Infrared (IR) sensor and image processing algorithm.

2. LITERATURE REVIEW

2.1. Works on Constructing Databases for SL Recognition

Usually, standard and complete signs video, point skeleton and depth stream database is connected with SLR for any research. Based on different SL in the world, researchers try to collect their own database with their own criteria and environmental condition to develop and perform SLR on their database.

The research in SL recognition has been conducted for different SLs in different countries [12-13, 15] and SL variants [5]. Many research in SL recognition only depicted the output of system and did not include the comprehensive dataset in the results. Several references that include the dataset applications are in [6, 9, 14], while in [15] presented two variants of the ASL (Greek and British).

Meanwhile, some detailed usage trends of an ASL database has been presented in [14] with explanations on data collection, organization, classification and retrieval. These data can provide the recognition algorithm developer with the opportunity to move from simple recognition situations in the best of circumstances to more complex recognition situations with challenging lighting situations. The video files are organized into alphabet, numbers 1-20, hand shapes with two samples for each signs, signs in isolation to show different motions, paragraphs to show connected discourse.

In [6] presented a paper shows ASL database included 2,576 images, which tried with 14 different native signers. His database was taken under certain condition such as lighting condition and high level studio. This database included static sign and dynamic signs. The American finger spelling alphabet, numbers, movement in single signs, examples of short discourse narratives for testing sign recognition in connected linguistic contexts were also included. In [5], the trend of organizing the SL dataset is using the search-by-example. The search-by-example allows the retrieving of database based on available example that is listed at the SL database.

From the review conducted, three important conclusions were derived:

1. Developing a SL dataset is one of the crucial parts of SL recognition. Many dataset variants of the SL exists and a MSL dataset does not yet exist.
2. The datasets reviewed consist of RGB images and videos, and no dataset has integrated skeletal data so far.
3. There is tremendous opportunity for research in SL recognition with the availability of a freely available online dataset.

2.2. Microsoft Kinect Operation

The Kinect device generates four outputs: an RGB image, an IR depth image, audio and skeleton joint locations based on its embedded sensors (RGB and depth cameras and microphone array) [10]. The IR receiver collects reflected IR beams transmitted from the IR transmitter. The IR transmitter emits infrared speckle patterns on its surrounding objects and the speckle patterns bounces off these objects. The reflected patterns are then received and processed by the IR receiver. The RGB sensor is a standard color camera capable of generating an image with a resolution of 640x480 pixels with 30 frames per second. Based on these two cameras, the Kinect algorithm can intelligently estimate the positions of 20 major joints in the body in skeleton form.

3. METHODOLOGY

This section describes the collection process and creation of the proposed MSL dataset. The following factors were considered for dataset creation:

1. Study the characteristic of MSL gestures and hand shapes.
2. Categorization of the signs based on one-handed and two-handed gestures, the position of the hand(s) with respect to the body, and hand direction.
3. Selection of signs which are the most important for daily communication.

Fig. 1 shows the process of developing the MSL dataset.

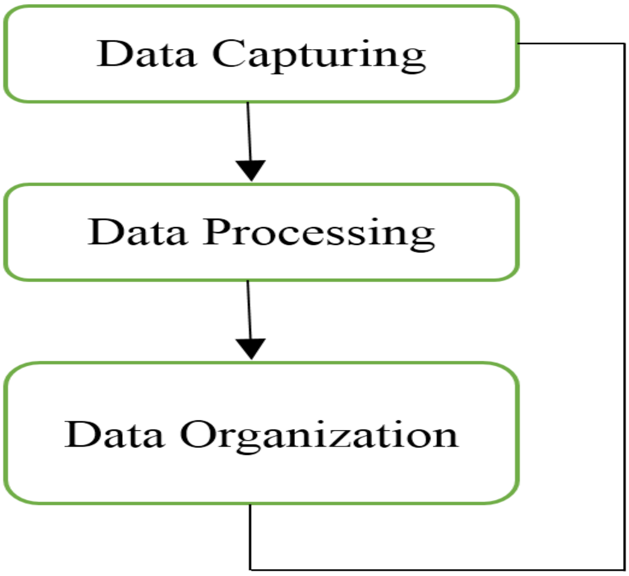


Fig.1. Process of developing MSL dataset

3.1. Data Capturing

The system uses the Microsoft Kinect sensor attached to a Personal Computer (PC). MATLAB 2016a was used as the Integrated Development Environment (IDE) to capture RGB, depth and skeleton data simultaneously. For the static signs (alphabets), the only hand locations were extracted and cropped. For dynamic signs, Kinect was used to detect joints in the upper body together with depth and RGB data.

Volunteer subjects to collect data were recruited from the Center of Deaf School Community who are fluent in MSL signing. The subjects were placed at a distance of 130 cm from the Kinect device. The recording process begins when the subject begins the hand motions. The subjects were instructed to perform the gestures at a controlled pace in order to ensure that the Kinect device can accurately capture the motions performed.

To obtain the proper SL data for the research, the data capture has the following conditions: (a) each recorded data contains an isolated sign, (b) and the participant has to repeat each sign five times and (c) all the recorded data should be kept.

Fig.2. Position of subject in front of camera

3.2. Data Processing

The captured data is processed to segment hand regions. We employ color, depth and skeleton joint data in order to extract hand segments.

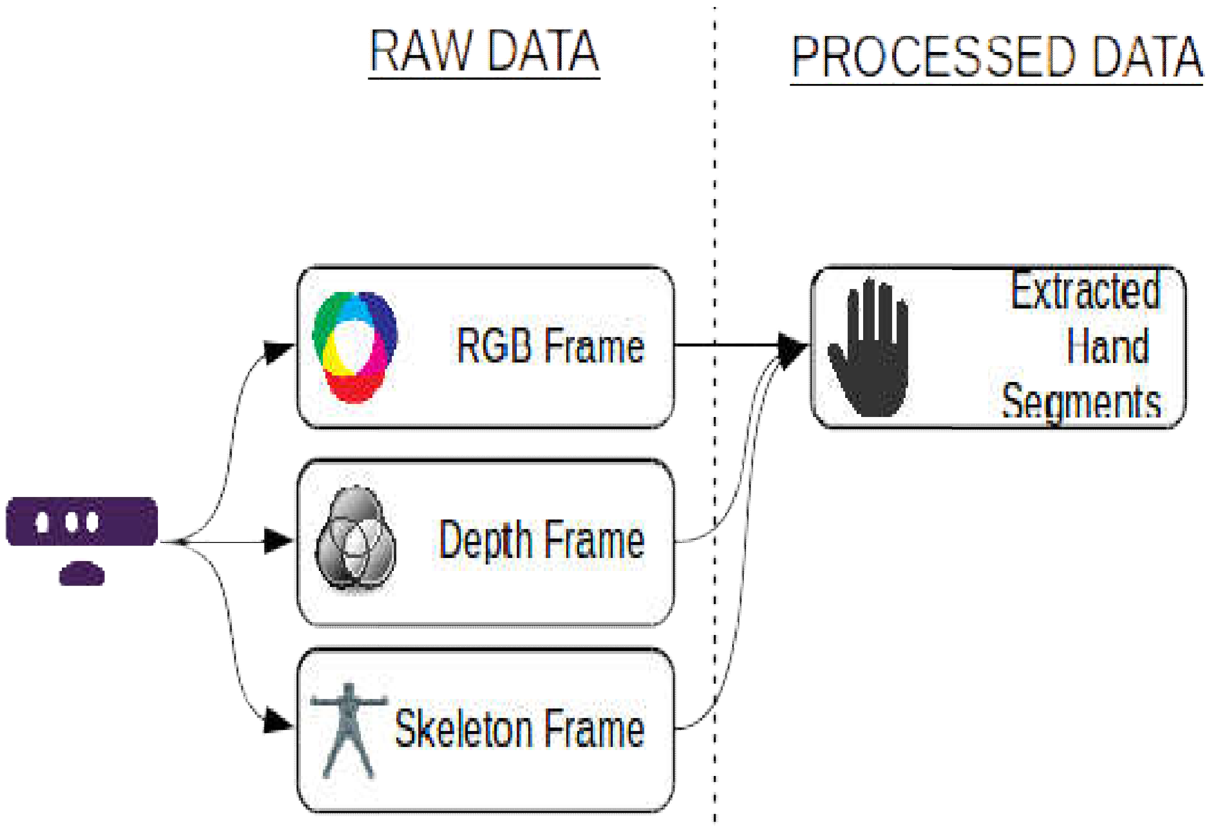


Fig.3. Data processing for extraction of static hand image (alphabets)

3.3. Data Organization

The recorded data are structured based on MSL dataset. Therefore, the recorded data are grouped based on these defined features:

1. Static signs (alphabets, 22 letters). All alphabets character except J and Z have been recorded. Letters J and Z need movement in the gesture to be present. So, they cannot be categorized as static signs.
2. Dynamic signs (10 signs). Most common dynamic signs which are similar to each other have been selected to be included in the dataset. Dynamic signs include signs which use right hand, left hand and both hands.

For dynamic sign, ten signs were chosen from a total of 42 candidate signs as shown in Table 1. The criteria for selecting the ten signs were similarity (we chose similar signs, which are difficult to discriminate) and difficulty (signs involving complex motions and hand shapes).

Table 1. Isolated word structures (Items in bold were chosen to be included in the dataset)

Grammar	Words							
Pronoun	**I**	**You**	They	She/he	We	**You**		
Verb	Learn	Walk	Sit	Talk	Stand	Eat	**Love**	Read
	Throw	**Divorce**	Drive	Wash	Bring	Lost	Hear	See
	Pray	Catch	Come	Put	**Want**	Know	Beat	Drink
Adjective	Beautiful	Big	Bad	Dirty	Good	Sick	-	-
Nouns	Car	Bus	Hand phone	**House**	Hospital	**God**	**Sister**	**Triangle**

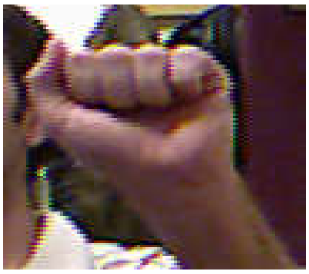
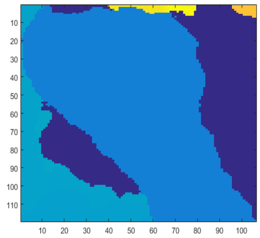
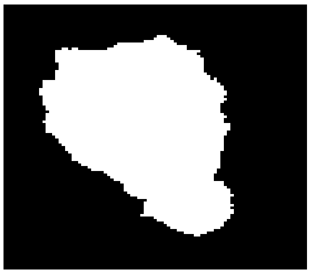
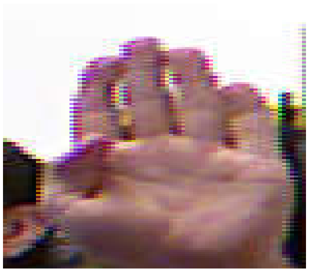
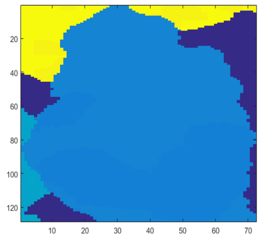
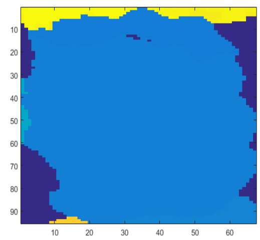
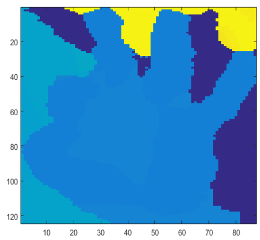
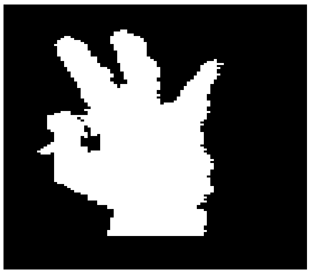
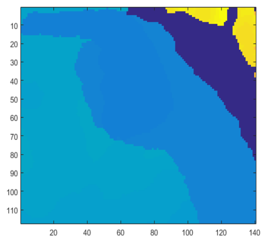
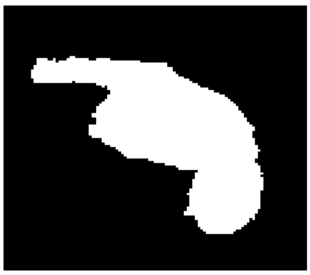
4. RESULTS AND DISCUSSION

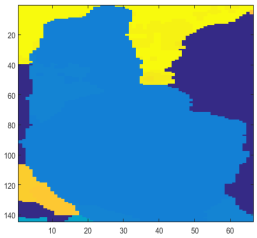
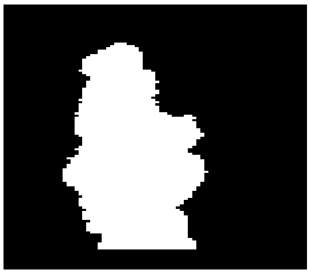
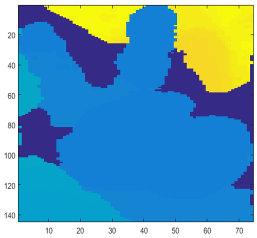
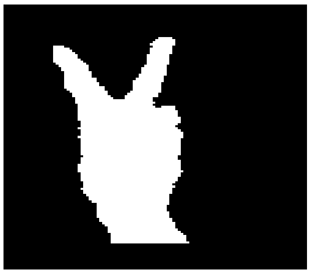
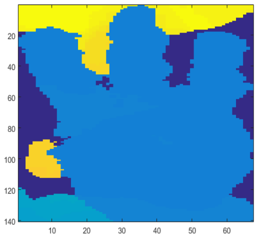
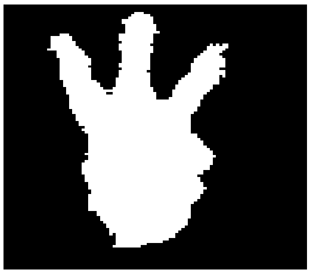
4.1. Static Signs (Alphabets)

The static signs consist of alphabets that do not require motion to represent. From a total of 24 alphabets, letters J and Z were removed as these two letters require motion to represent.

A sample of static signs collected is shown in Table 2.

Table 2. Samples of static signs collected

Characters	RGB	Depth	Threshold Method
A			
B			
C			
F			
G			

U			
V			
W			

As can be seen from Table 2, each letter is represented by a specific hand gesture. In the original RGB image, the image contains many background information which may interfere in the detection process. When depth image was used, the details in the background appears to be much reduced, which can bring a positive effect on classification. Additionally, using depth information, non-critical regions may be removed easily by setting a threshold that eliminates depth information beyond the threshold. This also can bring a positive result for classification [16, 18, 20-22]. An example of hand segmentation result is shown in Table 2.

4.2. Dynamic Signs

Dynamic signs require motion to convey the meaning of the SL. Selection of the dynamic signs take into account reasons outlined in section III-C. The signs include left-handed and right-handed signs together with signs that require both hands. Several examples of the dynamic signs are shown in Fig. 4 and Fig. 5.

Table 4. Sample frames for dynamic sign (divorce)

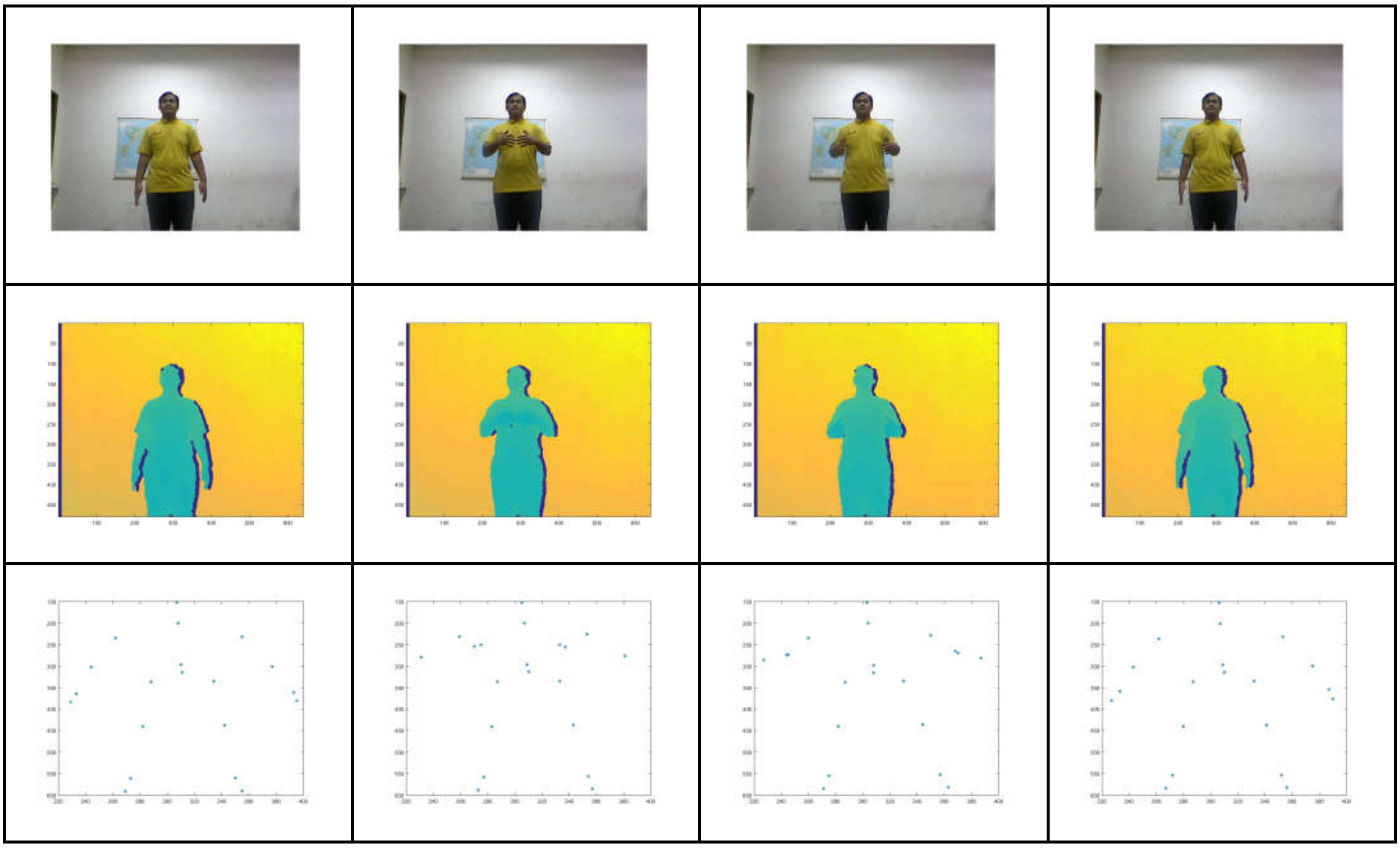


Fig5. Sample frames for dynamic sign (father)

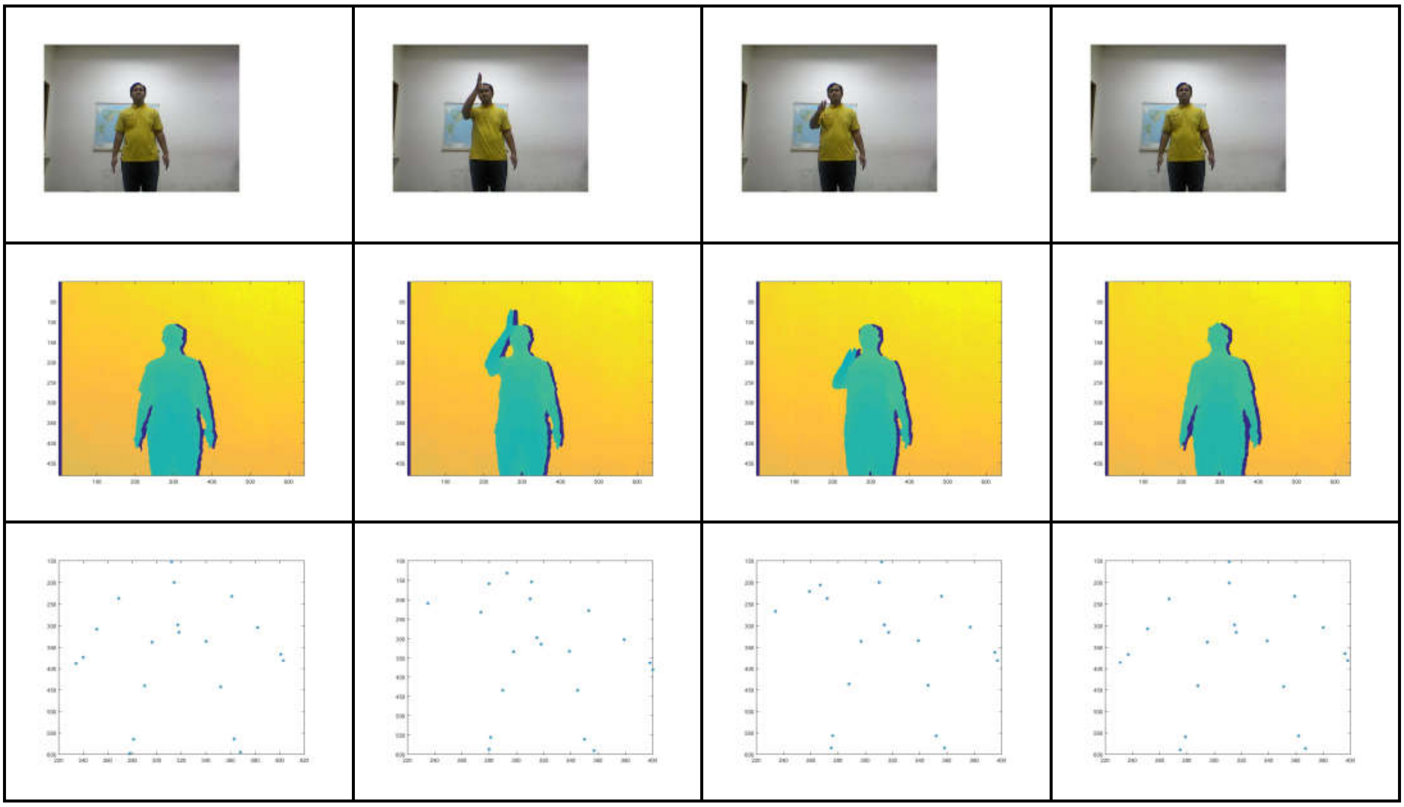


Fig.6. Sample frames for dynamic sign (sister)

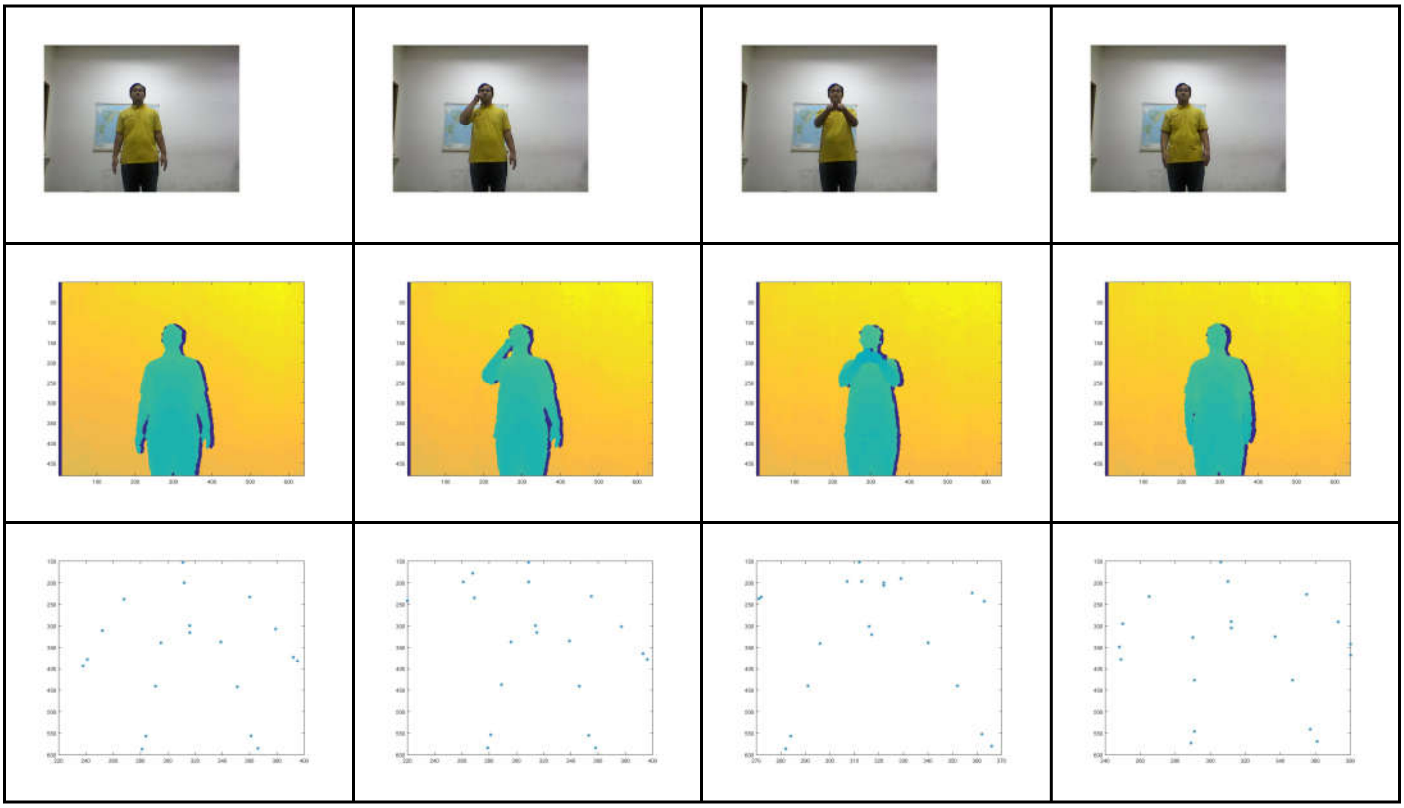
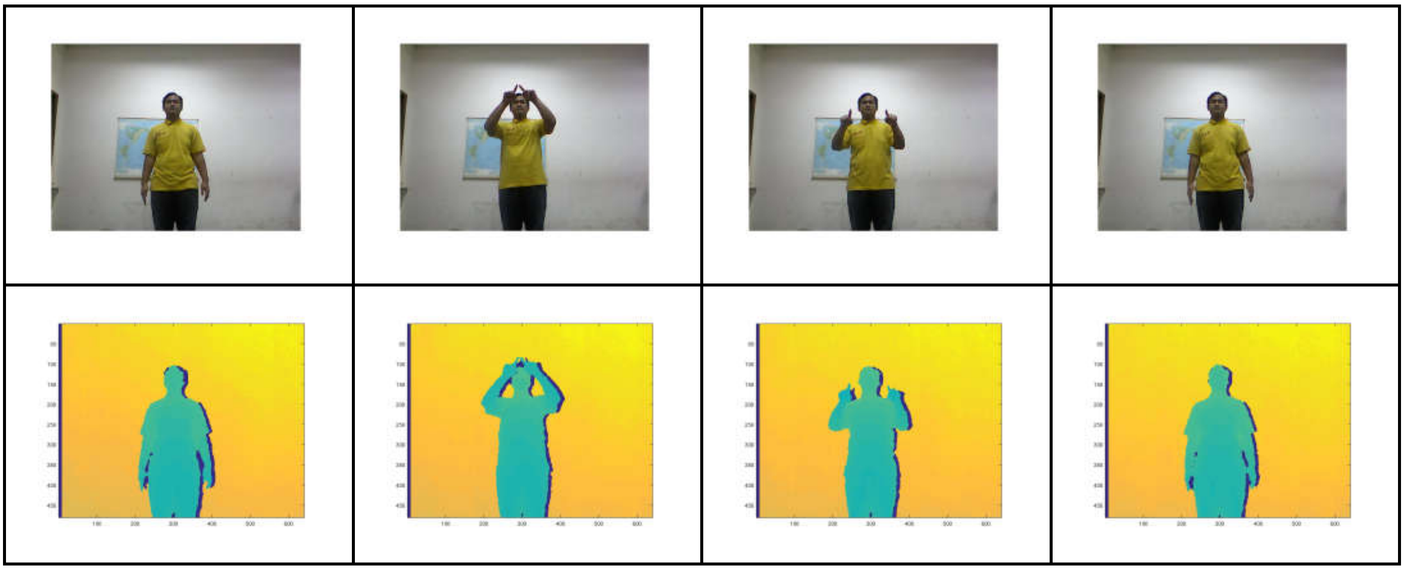
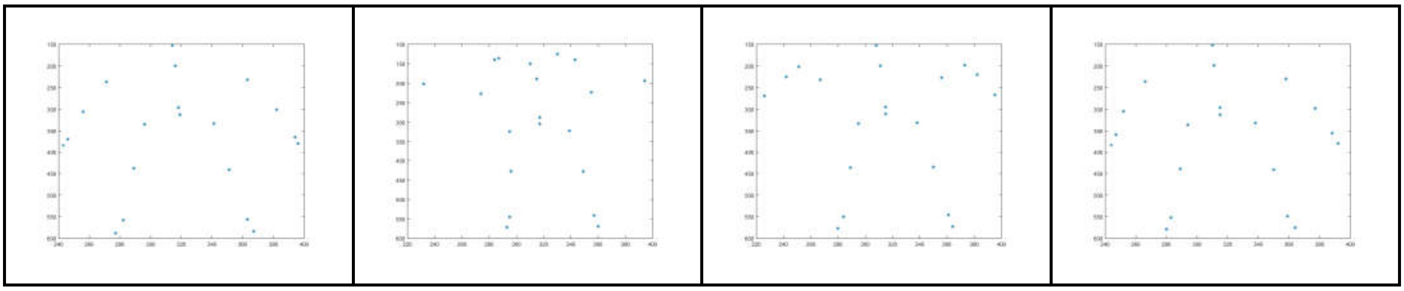


Fig.7. Sample frames for dynamic sign (triangle)





As can be seen from Fig. 6 and Fig. 7, each SL word require different gestures to perform. The gestures were able to be captured by Kinect in RGB, depth and joint skeleton. Each of the features were able to represent the different gestures involved. The depth information was able to eliminate many background objects compared to RGB camera (map picture on the wall), showing its potential for extraction of regions of interest while removing non-essential features seamlessly. The skeleton image depicts the lower body points even though this does not appear in the RGB or depth image. This is because the lower body points below the knee are extrapolated automatically by the Kinect device based on the detected hip and knee joint positions. The extrapolated points have a tendency to be inaccurate. However, for our application, this is negligible because we only use the upper part of the body for sign language detection. The point that were selected for this purpose is Fig. 8.

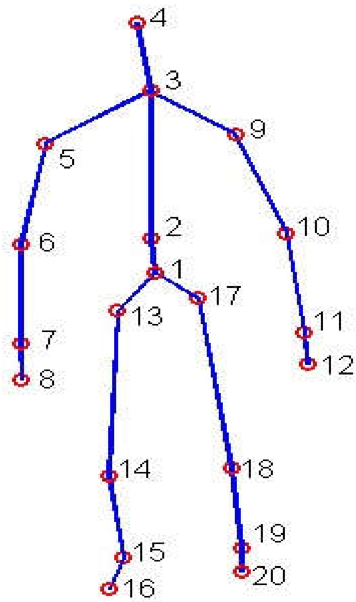


Fig.8. Source (internet)

5. CONCLUSION

The development of a MSL dataset using Microsoft Kinect for automated SL recognition is presented in this paper. We describe the implementation and collection of data in detail. The use of Microsoft Kinect was able to include a much richer feature set, which we believe would be able to significantly improve the recognition accuracy compared to current methods.

6. ACKNOWLEDGEMENTS

The authors would like to express their gratitude to International Islamic University Malaysia and Universiti Teknologi MARA for the equipment and infrastructure in support of this research.

7. REFERENCES

[1] Abdullah M S, Zabidi A, Yassin I M, Hassan H A. Analysis of Microsoft Kinect depth perception for distance detection of vehicles. In IEEE 6th Control and System Graduate Research Colloquium, 2015, pp. 116-119

[2] Abiddin W Z, Jailani R, Omar A R, Yassin I M. Development of MATLAB Kinect Skeletal Tracking System (MKSTS) for gait analysis. In IEEE Symposium on Computer Applications & Industrial Electronics, 2016, pp. 216-220

[3] Akmeliawati R, Ooi M P, Kuang Y C. Real-time Malaysian sign language translation using colour segmentation and neural network. In IEEE Instrumentation and Measurement Technology Conference, 2007, pp. 1-6

[4] Laws of Malaysia. Federal constitution. 2006, http://defensewiki.ibj.org/images/c/cc/Malaysian_Constitution.pdf

[5] Elliott R, Cooper H, Ong E J, Glauert J, Bowden R, Lefebvre-Albaret F. Search-by-example in multilingual sign language databases. In Sign Language Translation and Avatar Technologies Workshops, 2011, pp. 1-8.

[6] Kak A C. Purdue RVL-SLLL ASL database for automatic recognition of American Sign Language. In 4th IEEE International Conference on Multimodal Interfaces 2002, pp. 1-6

[7] Karabasi M, Bhatti Z, Shah A. A model for real-time recognition and textual representation of Malaysian sign language through image processing. In IEEE International Conference on Advanced Computer Science Applications and Technologies, 2013, pp. 195-200

[8] Ladner R. Technology for deaf people. 2009, https://pdfs.semanticscholar.org/8462/b3f9411335bee9be5548141c18a7f83623e8.pdf

[9] Maarif H A, Akmeliawati R, Bilal S. Malaysian Sign Language database for research. In IEEE International Conference on Computer and Communication Engineering, 2012, pp. 798-801

[10] Manap M S, Sahak R, Zabidi A, Yassin I, Tahir N M. Object detection using depth information from Kinect sensor. In IEEE 11th International Colloquium on Signal Processing and Its Applications, 2015, pp. 160-163

[11] Ong S C, Ranganath S. Automatic sign language analysis: A survey and the future beyond lexical meaning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 1(6):873-891

[12] Wang C, Chen X, Gao W. Expanding training set for Chinese sign language recognition. In 7th IEEE International Conference on Automatic Face and Gesture Recognition, 2006, pp. 323-328

[13] Werapan W, Chotikakamthorn N. Improved dynamic gesture segmentation for Thai sign language translation. In 7th IEEE International Conference on Signal Processing, 2004, pp. 1463-1466

[14] Wilbur R, Kak A C. Purdue RVL-SLLL American sign language database. ECE Technical Reports, Indiana: Purdue University, 2006

[15] Zhang L G, Chen Y, Fang G, Chen X, Gao W. A vision-based sign language recognition system using tied-mixture density HMM. In ACM 6th international conference on Multimodal interfaces 2004, pp. 198-204

[16] Zaman F H, Sulaiman A A, Yassin I M, Tahir N M, Rizman Z I. Development of mobile face verification based on locally normalized Gabor wavelets. International Journal on Advanced Science, Engineering and Information Technology, 2017, 7(4):1198-1205

[17] Zabidi A, Tahir N M, Yassin I M, Rizman Z I. The performance of binary artificial bee colony (BABC) in structure selection of polynomial NARX and NARMAX models. International Journal on Advanced Science, Engineering and Information Technology, 2017, 7(2):373-379

[18] Latip M F, Udin M K, Othman M M, Yassin I M, Rizman Z I, Zaini N, Hidayat M N, Aminuddin N, Herman S H, Saad H, Rahiman M H. Implementation of fuzzy logic-based final year project student-supervisor matching system. International Journal of Advanced and Applied Sciences, 2017, 4(4):159-163

[19] Yassin I M, Jailani R, Ali M S, Baharom R, Hassan A H, Rizman Z I. Comparison between cascade forward and multi-layer perceptron neural networks for NARX functional electrical stimulation (FES)-based muscle model. International Journal on Advanced Science, Engineering and Information Technology, 2017, 7(1):215-221

[20] Yassin I M, Zan M M, Tahir N M, Abidin H Z, Rizman Z I. Integration of CompTIA Cloud+ into Universiti Teknologi MARA's computer engineering special topics syllabus. International Journal on Advanced Science, Engineering and Information Technology, 2017, 7(1):22-27

[21] Hasan C Z, Jailani R, Tahir N M, Yassin I M, Rizman Z I. Automated classification of autism spectrum disorders gait patterns using discriminant analysis based on kinematic and kinetic gait features. Journal of Applied Environmental and Biological Sciences, 2017, 7(1):150-156

[22] Nor M M, Jailani R, Tahir N M, Yassin I M, Rizman Z I, Hidayat R. EMG signals analysis of BF and RF muscles in autism spectrum disorder (ASD) during walking. International Journal on Advanced Science, Engineering and Information Technology, 2016, 6(5):793-798

[23] Yassin I M, Zabidi A, Ali M S, Tahir N M, Abidin H Z, Rizman Z I. Binary particle swarm optimization structure selection of nonlinear autoregressive moving average with exogenous inputs (NARMAX) model of a flexible robot arm. International Journal on Advanced Science, Engineering and Information Technology, 2016, 6(5):630-637.

How to cite this article:
Karbasi M, Zabidi A, Yassin IM, Waqas A, Bhatti Z. Malaysian sign language dataset for automatic sign language recognition system. J. Fundam. Appl. Sci., 2017, 9*(4S), 459-474*.