

# A REVIEW ON THE DEVELOPMENT OF INDONESIAN SIGN LANGUAGE RECOGNITION SYSTEM

<sup>1,2</sup>Sutarman, <sup>1</sup>Mazlina Abdul Majid and <sup>1</sup>Jasni Mohamad Zain

<sup>1</sup>Department of Software Engineering, Universiti Malaysia Pahang, Kuantan, Pahang, Malaysia

<sup>2</sup>Department of Information Technology and Business, University Technology of Yogyakarta, Indonesia

Received 2013-06-02, Revised 2013-07-23; Accepted 2013-09-27

## ABSTRACT

Sign language is mainly employed by hearing-impaired people to communicate with each other. However, communication with normal people is a major handicap for them since normal people do not understand their sign language. Sign language recognition is needed for realizing a human oriented interactive system that can perform an interaction like normal communication. Sign language recognition basically uses two approaches: (1) computer vision-based gesture recognition, in which a camera is used as input and videos are captured in the form of video files stored before being processed using image processing; (2) approach based on sensor data, which is done by using a series of sensors that are integrated with gloves to get the motion features finger grooves and hand movements. Different of sign languages exist around the world, each with its own vocabulary and gestures. Some examples are American Sign Language (ASL), Chinese Sign Language (CSL), British Sign Language (BSL), Indonesian Sign Language (ISL) and so on. The structure of Indonesian Sign Language (ISL) is different from the sign language of other countries, in that words can be formed from the prefix and or suffix. In order to improve recognition accuracy, researchers use methods, such as the hidden Markov model, artificial neural networks and dynamic time warping. Effective algorithms for segmentation, matching the classification and pattern recognition have evolved. The main objective of this study is to review the sign language recognition methods in order to choose the best method for developing the Indonesian sign language recognition system.

**Keywords:** Indonesian Sign Language, Recognition System, Hidden Markov Model, Artificial Neural Network, Dynamic Time Warping

## 1. INTRODUCTION

Sign language is one of the most important and natural communication modalities. It is a static expression system that is composed of signs by using hand motion aided by facial expressions. Sign language is mainly employed by hearing-impaired people to communicate with each other. However, communication with normal people is a major handicap for them since normal people do not understand their sign language.

Sign language recognition is needed for realizing a human oriented interactive system, which can perform an interaction like normal communication. Without a translator, most people who are not familiar

with sign language will have difficulty to communicate. Thus, software that transcribes symbols in sign languages into plain text can help with real-time communication and may also provide interactive training for people to learn a sign language.

Gesture recognition has become an important research field with the current focus on interactive emotion recognition and hand gesture recognition. Kinect, as an XBOX motion sense game controller, can be used to obtain to garner video with depth information as well as track the skeletal movements of the gamer. Now the Kinect can also be used to recognize body motion without connecting to an Xbox because it can be connected to a normal PC to collect information.

**Corresponding Author:** Sutarman, Department of Software Engineering, Universiti Malaysia Pahang, Kuantan, Pahang, Malaysia

Research in the field of sign language can be categorized into two; one that employs a vision-based computer (computer vision), while the other is based on the sensor data (Ma *et al.*, 2000; Maraqa and Abu-Zaiter, 2008). In computer vision-based gesture recognition, a camera is used as input. Videos are captured in the form of video files stored before being processed using image processing (Brashear *et al.*, 2003; Hninn and Maung, 2009; Stergiopoulou and Papamarkos, 2009).

Several works on sign language system have been previously proposed, which mostly make use of probabilistic models, such as Hidden Markov Models and Artificial Neural Networks. Thus, the main objective of this study is to review and compare the performance of sign language recognition methods. The purpose of the comparison is to find out the best sign recognition method for developing the Indonesian Sign Language (ISL).

## 2. A REVIEW OF THE SIGN LANGUAGE SYSTEM

### 2.1. Sign Language

The term sign language is similar to the language term, in that there are many of both spread throughout the various territories of the world. Just like other languages, sign language was developed a long time ago and has developed over a long time, has a sign language grammar and vocabulary and, hence, is considered a real language (Braem, 1995).

This language is commonly used in deaf communities, including by interpreters, friends and families of the deaf, as well as people who are hard of hearing themselves. However, these languages are not commonly known outside of these communities and therefore, communication barriers exist between deaf and hearing people.

Sign language communication is multimodal. It involves both hand gestures (i.e., manual signing) as well as non-manual signals. Gestures in sign language are defined as specific patterns or movements of the hands, face or body to make our expressions.

Since it is a natural language, sign language is closely linked with the culture of the deaf, from which it originates. Thus, knowledge of the culture is necessary to fully understand sign language.

The difference between sign language and common language concerns the method to communicate/articulate information. Because no sense of hearing is required to understand sign language and no voice is required to

produce it, it is the common type of language among deaf people (Lang *et al.*, 2011).

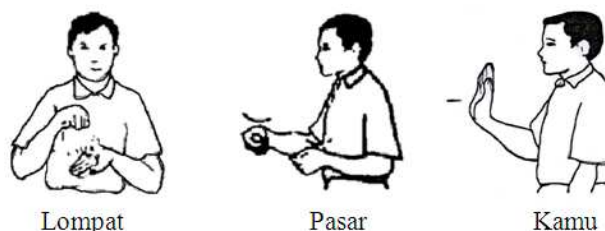
### 2.2. Indonesian Sign Language (ISL)

The Indonesian Sign Language (ISL), which is also known as Sistem Isyarat Bahasa Indonesia (SIBI), can be broadly divided into two, namely, the alphabet gesture and the word gesture. For the alphabet gesture, SIBI refers to the American Sign Language (ASL), while for the word gesture the word in Indonesian is used to symbolize its common meaning. The alphabet gesture is usually used to spell names or words that are not listed in the dictionary of the vocabulary. Word gestures are more widely used in practice and have a much larger sign. Both gesture signs have components to the gesture. The main component to the signal is formed by the fingers and hand movements. The word gesture frequently uses a hand gesture, instead of a finger formation.

Types of gesture in ISL include: (a) basic gesture (the basic word) as depicted as **Fig. 1**, (b) additional gesture (prefix, suffix) as depicted in **Fig. 2**, (c) formation of the gesture (combined gesture) as depicted in **Fig. 3**, (d) finger alphabet (Character/Numeric) as depicted in **Fig. 4**.

Dictionary ISL (DPN, 2001) has been accepted by the Ministry of Education of Indonesia to be used as the national standard. ISL standardized it as one of the media to help communication among the deaf within the larger society. ISL consists of systematic rules for gestures of fingers, hands and other movements that symbolize the Indonesian vocabulary. This standard was accepted based on some considerations, such as easiness, properness and accuracy of the meanings and structure of the language.

The research on Indonesian sign language is still very little and still need development. The research is still largely in the form of applications, while for studies using recognition methods such as HMM, ANN, DTW is still little and results accuracy needs to be improved.



**Fig. 1.** Example of basic gestures

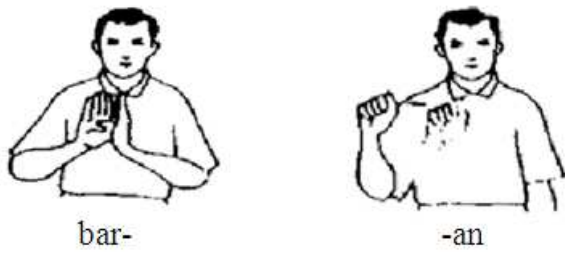


Fig. 2. Examples additional gestures (prefix, suffix)

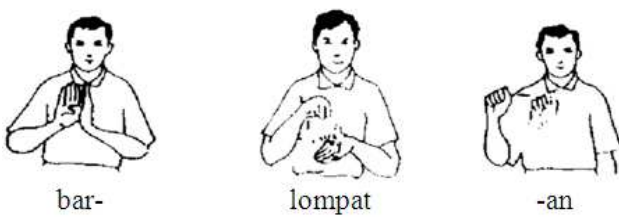


Fig. 3. Examples of combined gesture (berlompatan)

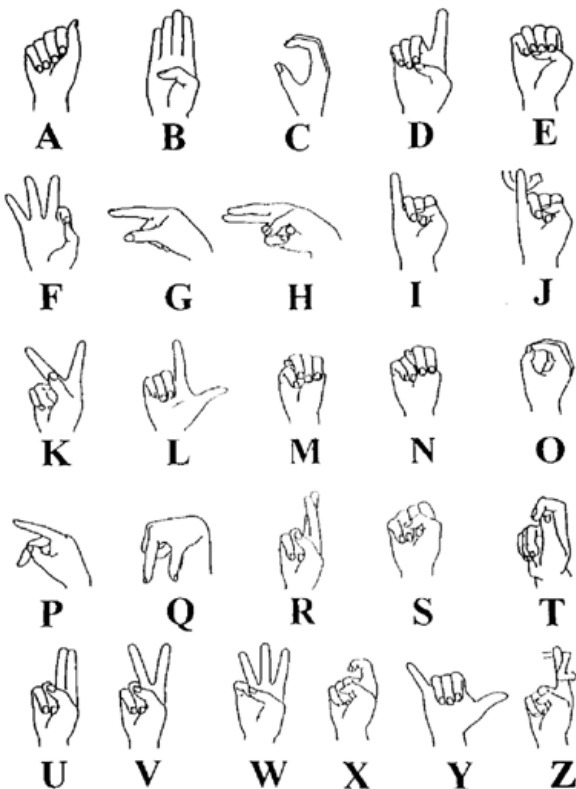


Fig. 4. Examples of finger alphabet (Charater/Numeric)

Research objects with SIBI such as: The Indonesian Sign Language for the Deaf Computer Application developed by Yanuardi *et al.* (2010) and Asriani and Susilawati (2010) used the Backpropagation Neural Networks (BNN) method for the Indonesian Sign Language (ISL) recognition, Sekar (2001) used the Neural Networks (NN) method for the Indonesian Sign Language (ISL), Iqbal *et al.* (2011) used DTW for the Indonesian Sign Language based-sensor accelerometer and sensor flex.

### 3. SIGN CAPTURING METHODS

#### 3.1. Data Glove Approach

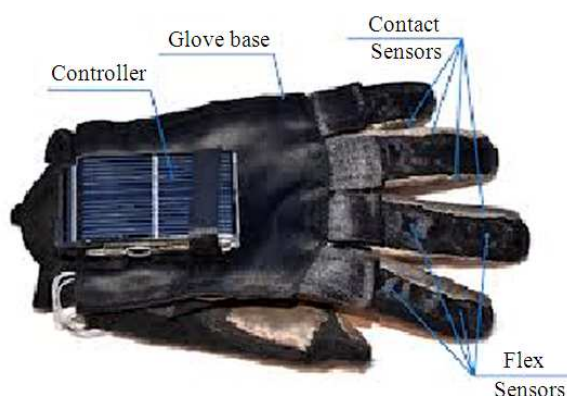
One of the traditional sign capturing method is data glove approach (Fig. 5). These methods employ mechanical or optical sensors attached to a glove that transforms finger flexions into electrical signals to determine the hand posture (Xingyan, 2003). This method requires the glove must be worn and a wearisome device with a load of cables connected to the computer, which will hamper the naturalness of user-computer interaction. However, the disadvantage of this method is it requires the glove to be worn, which is a wearisome device with many cables connected into the computer that will hamper the naturalness of the user computer interaction (Mitra and Acharya, 2007).

#### 3.2. Vision-based Sign Extraction

Another common sign capturing method is vision-based sign extraction. This method is usually done by capturing an input image using a camera (Fig. 6). In order to create a database for a gesture system, the gestures should be selected with their relevant meaning, in which each gesture may contain multi samples (Hasan and Mishra, 2010) to increase the accuracy of the system.

Vision-based method is widely deployed for sign language recognition. Sign gestures are captured by a fixed camera in front of signers. The extracted images convey posture, location and motion features of the fingers, palms and face. Next, an image-processing step is required in which each video frame is processed in order to isolate the signer's hands from other objects in the background.

The problem is that errors relate to a dynamic environment. Furthermore, the vast computation needed is another issue with a real time vision system. For example, ElenaSanchez-Nielsen *et al.* (2003) suggested a real time vision system that uses a fast segmentation method, by using minimum features to identify hand posture in order to speed up the recognition process.



**Fig. 5.** Example of the data glove



**Fig. 6.** Examples of the data vision

Brashear *et al.* (2003) used a camera mounted above the signer, so the images captured by this camera clearly solve the overlapping between the signer's hands and head. However, unfortunately, face and body gestures are lost this way.

### 3.3. Microsoft Kinect XBOX 360™

Microsoft Kinect provides an inexpensive and easy way for real-time user interaction. The software driver released by Microsoft called Kinect Software Development Kit (SDK) with Application Programming Interfaces (API) gives access to raw sensor data streams as well as skeletal tracking (Zhang *et al.*, 2012). Although there is no hand specific data available for gesture recognition, it does include information of the joints between hands and arms. Little work has been done for Kinect to detect the details of the level of individual fingers. **Figure 7** is the example of Kinect device.



**Fig. 7.** Kinect device

Framework for recognition using a Kinect created by Lang *et al.* (2011) called the Dragonfly movement-Draw on the fly and mainly consists of two classes that can be used within the assimilation of users with their software. The depth of the camera serves as an interface for OpenNI, i.e., it updates the camera image and report data of the skeleton joints of the body and the final data set for recognition process.

The Kinect and depth cameras, in general, are well-suited for sign language recognition. They offer 3D data from the environment without a complicated camera setup and efficiently extract the users' body parts, allowing for recognition of not just hands and head, but also other parts such as elbows that can be of further help in distinguishing between similar signs. Another advantage is the independency of lighting conditions, as the camera uses infrared light. All body parts are detected equally well in a dark environment and there is no need from the user to wear special colored gloves or wired gloves.

## 4. SIGN LANGUAGE RECOGNITION METHODS

Many methods have been used for sign language recognition, some of which are described in the following subsections.

### 4.1. Hidden Markov Model (HMM)

Hidden Markov Models (HMMs) are learnable finite stochastic automates. They are considered as a specific form of dynamic Bayesian networks. A Hidden Markov Model consists of two stochastic processes. The first stochastic process is a Markov chain that is characterized



by states and transition probabilities. The states of the chain are externally not visible, therefore “hidden”. The second stochastic process produces emissions observable at each moment, depending on a state-dependent probability distribution (Dymarski, 2011).

HMM is used in robot movement, bioinformatics, speech and gesture recognition. This model has two advantages regarding sign recognition, the ability to model linguistic roles and its ability to classify continuous gestures within a certain assumption (Ng *et al.*, 2002).

Septiari and Haryanto (2012) used speech recognition methods Hidden Markov Model (HMM) for the Indonesian Sign Language. They have conducted two experiments with unsatisfactory results. The main obstacle in doing this research is the introduction of the noise of the system to text while conversion from text to video sign language did not experience significant obstacles in terms of programming.

Xiaoyu *et al.* (2008) presented multilayer architecture in sign language recognition for the signer-independent CSL recognition, in which classical Dynamic Time Warping (DTW) and HMM are combined within an initiative scheme. In the two-stage hierarchy, they define the confusion sets and introduce the DTW/ISODATA algorithm as the solution to build confusion sets in the vocabulary space. The experiments show that the multilayer architecture in sign language recognition increases the average recognition time by 94.2% and the recognition accuracy 4.66% more than the HMM-based recognition method.

Sandjaja and Marcos (2009) used the HMM for the training and testing phase in Filipino sign language number. The feature extraction could track 92.3% of all objects. The recognizer could also recognize Filipino Sign Language numbers with an average of 85.52% accuracy.

Lang *et al.* (2011) used the HMM for sign language recognition using Kinect. The sign language recognition framework makes use of Kinect, a depth camera developed by Microsoft and PrimeSense, which features easy extraction of important body parts. The framework also offers an easy way of initializing and training new gestures or signs by performing them several times in front of the camera. The results show a recognition rate of >97.00% for eight out of nine signs when they are trained by more than one person.

Bowden *et al.* (2004) used Markov chains in combination with Independent Component Analysis

(ICA) for a system for recognizing the British Sign Language (BSL). It captures data using an image technique then extracts a feature set describing the location, motion and shape of the hands based on BSL sign linguistics. High level linguistic features were used to reduce the recognizer’s training work. Classification rates as high as 97.67% were achieved for a lexicon of 43 words using only single instance training.

When analyzing the HMM method, the best result recognition and accuracy are: (a). Lang *et al.* (2011) for sign language recognition using Kinect, with results recognition rate of >97.00% for eight out of nine signs; (b). Bowden *et al.* (2004) used Markov chains in combination with Independent Component Analysis (ICA) for a system for recognizing British Sign Language (BSL). The classification rates were as high as 97.67% for a lexicon of 43 words using only single instance training.

HMMs are strong in its learning ability which is achieved by presenting time-sequential data and automatically optimizing the model with the data. For HMMs, Baum-Welch algorithm is the most frequently used training method. Although Baum-Welch algorithm is a very efficient tool for training, its weakness is that it is easy to get into the local optimums due to its defects (Zhang *et al.*, 2012).

The conclusion is that to be able to produce this level of accuracy and classification, it is necessary to combine methods, such as HMM with ICA (Bowden *et al.*, 2004) and also use data capture technologies, such as Kinect (Lang *et al.*, 2011).

## 4.2. Artificial Neural Networks (ANN)

Many researchers highlight the success of using neural networks in sign language recognition. An Artificial Neural Network (ANN) consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation.

Hasan and Mishra (2010) used a Multilayer Perceptron (MLP) neural network to recognize static hand-finger gestures of the yubimoji, the Japanese Sign Language syllabary. Signal inputs from the data glove interface were taken separately for each static yubimoji gesture. Each input was fed as input of MLP, after which the network was trained 10 times and tested for 41 gestures. Generally, only 18 of the static gestures were successfully recognized. One of the reasons was attributed to the data glove’s inability to measure gesture

directions, particularly yubimoji gestures with similar finger configurations.

Karami *et al.* (2011) used Wavelet transform and Neural Networks (NN) for Persian Sign Language (PSL) recognition. The system was implemented and tested using a data set of 640 samples of Persian sign images; 20 images for each sign. The experimental results show that the system can recognize 32 selected PSL alphabets with an average classification accuracy of 94.06%.

Maraqa and Abu-Zaiter (2008) used two recurrent neural networks architectures for static hand gestures to recognize the Arabic Sign Language (ArSL). Elman (partially) recurrent neural networks and fully recurrent neural networks were used. A digital camera and a colored glove were used for input image data. For the segmentation process, the HSI color model was used. Segmentation divides the image into six color layers, five for fingertips and one for the wrist. Thirty features are extracted and grouped to represent a single image, expressed by the fingertips and the wrist with angles and distances between them. This input feature vector is the input to both neural network systems. A total of 900 colored images were used for the training set and 300 colored images for testing purposes. The results showed that the fully recurrent neural network system (with recognition rate 95.11%) is better than the Elman neural network (89.67%).

Admasu and Raimond (2010) used the Gabor Filter (GF) together with Principal Component Analysis (PCA) for extracting features from the digital images of hand gestures for the Ethiopian Sign Language (ESL), while the Artificial Neural Network (ANN) was used for recognizing the ESL from extracted features and translation into Amharic voice. The experimental results show that the system produced a recognition rate of 98.53%.

Hninn and Maung (2009) used real time 2D hand tracking to recognize hand gestures for the Myanmar Alphabet Language. Digitized photograph images were used as input images and the Adobe Photoshop filter was applied for finding the edges of the image. By employing histograms of local orientation, this orientation histogram was used as a feature vector. MATLAB toolbox was used for system implementation. Experiments show that the system can achieve a 90.00% recognition average rate.

Bailador *et al.* (2007) presented a Continuous Time Recurrent Neural Networks (CTRNN) real time hand

gesture recognition system using a tri-axial accelerometer sensor and wireless mouse to capture the 8 gestures used. The work was based on the idea of creating specialized signal predictors for each gesture class, in which standard a Genetic Algorithm (GA) was used to represent the neuron parameters. Each genetic string represents the parameter of a CTRNN. Two datasets were applied; one for isolated gestures, with a recognition rate of 98% for the training set and 94% for the testing set. For the second dataset, for captured gestures in a real environment, for the first set, the recognition rate was 80.50% for training and 63.60% for testing.

Stergiopoulou and Papamarkos (2009) conducted a study on the static hand gesture recognition based Neural Gas Self-Growing and Self-Organized (SGONG) network. An input image using a digital camera for the detection of the hand area of YCbCr color space was applied and the threshold technique was used to detect skin tones. They uses the competitive Hebbian learning algorithm, which begins studying with two neurons. As the neurons grow the grid will detect the exact shape of the hand, with the specified number of fingers raised, however, in some cases the algorithm might lead to false classification. This problem is solved by applying the average finger length ratio. This method has the disadvantage that two fingers may be classified into the same class of the finger. This problem has been overcome by choosing the most likely combinations of fingers. This system can recognize the 31 movements that have been established with a recognition rate of 90.45% and 1.5 sec.

Akmeliawati *et al.* (2007) presented an automatic visual-based sign language translation system. They proposed automatic sign-language translator provides a real-time English translation of the Malaysian SL. The sign language translator can recognize both finger spelling and sign gestures that involve static and motion signs. Its trained neural networks are used to identify the signs to translate into English.

Zhang *et al.* (2012) presented a method of scoring time sequential postures of golf swing Classification System Using HMM and Neuro-Fuzzy. The results show that the proposed methods can be implemented to identify and score the golf swing effectively with up to 80.00% accuracy.

Al-Jarrah and Halawani (2001) developed a system for the automatic translation of gestures of the manual alphabets in the Arabic Sign Language. The system recognized the 30 Arabic Sign Language alphabets

visually, using images of the bare hands. They used the Adaptive Neuro-Fuzzy Inference system (ANFIS) to accomplish the recognition job. The success rate of recognition accuracy was 93.55%.

Binh and Ejima (2005) proposed a new approach to hand gestures recognition based on feature recognition neural network in which develop a neural network architecture, which incorporates the idea of fuzzy ARTMAP in feature recognition neural network. Experiments show that the system can achieve a 92.19% recognition average rate.

Asriani and Susilawati (2010) used the Backpropagation Neural Networks (BNN) method for the Indonesian Sign Language (ISL) recognition. The success rate for static hand gesture recognition achieved in this study was 69%.

Sekar (2001) used the Neural Networks (NN) method for the Indonesian Sign Language (ISL) recognition, in which the data processed, is obtained from the sensor flex grooves information covering the fingers, wrists curve, the curve of the arm and shoulder grooves. The dataset comprises 72 cue words SIBI static. Results: 83.18% for 22 words (only using fingers + wrist) and 49.58% for 72 words (using all sensors).

### 4.3. Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) was introduced in the 1960s (Bellman and Kalaba, 1959). It is an algorithm for measuring the similarity between two sequences, which may vary in time or speed. For instance, similarities in walking patterns would be detected, even if in one video, the person is walking slowly and if in another video, he or she is walking more quickly, or even if there are accelerations and decelerations during one observation.

DTW has been used in video, audio and graphics applications. In fact, any data that can be turned into a linear representation can be analyzed with DTW (a well-known application has been automatic speech recognition). By using DTW, a computer is able to find an optimal match between two given sequences (i.e., signs) with certain restrictions. The sequences are "warped" non-linearly in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension.

Li and Greenspan (2007), using Compound Gesture Models, in which the temporal endpoints of a gesture were estimated by DTW and a bounded search was performed to recognize the gesture. The proposed method is both computationally efficient and robust. In

experiments containing nine different gestures and five subjects, the resulting average recognition rates were 93,30% for single scale and 88,10% for multiple scale continuous gestures.

The Microsoft Kinect XBOX 360™ is proposed to solve the problem of sign language translation (Capilla, 2012). By using the tracking capability of this RGB-D camera, a meaningful 8-dimensional descriptor for every frame is introduced here. In addition, an efficient Nearest Neighbor DTW and Nearest Group DTW are developed for fast comparison between sign languages. For a dictionary of 14 homemade signs, the introduced system achieves an accuracy of 95.24%.

Iqbal *et al.* (2011) used DTW for the Indonesian Sign Language based-sensor accelerometer and sensor flex. The experiments were conducted to recognize the 50-word (classes) of Sistem Isyarat Bahasa Indonesia (SIBI). The selected words are only implied by one hand, i.e., the right hand. The results show that the DTW method used can recognize words with an average accuracy of 95.60%.

The DTW-algorithm is used to compare two signs no matter their length. By doing so, the system is able to deal with different speeds during the execution of two different samples for the same sign, but sometimes the algorithm can wrongly output a positive similarity coefficient.

### 4.4. Summary Comparisons of HMM, ANN and DTW Methods

**Table 1** shows the summary of comparison for HMM, ANN and DTW according to the discussion in above.

A comparison on the advantages and disadvantages of sign language recognition methods for ANN, DTW and HMM was made between each of these methods. It is found that different ANN systems are used in different stages of recognition systems according to the nature of the problem, its complexity and the environment available. Additionally, since true human gestures are continuous, introducing an isolated system can significantly disrupt the natural flow of human interaction and it does not have as much value in the reality of sign recognition. The success of a fully automated sign recognition system relies on solving current problems associated with continuous gesture recognition. HMM classifier also proves interesting in sign recognition due its ability to model words based on sets of predefined states.

**Table 1.** The literature comparison for HMM, ANN and DTW methods

Author	Method	Recognition Rate	Recognition Limited
Lang <i>et al.</i> (2011) for Sign Language Recognition Using Kinect	HMM	Recognition rate of > 97.00%	The experiment for eight out of nine signs
Capilla (2012). For sign language translator using Microsoft Kinect XBOX360™	DTW	The result accuracy of 95.24%.	The project does not focus on particular official dictionary of signs
Zhang <i>et al.</i> (2012) for postures of golf swing Classification System	HMM and neuro-fuzzy	Score the golf swing effectively with up to 80.00% accuracy	the precision of Joint position captured by Kinect was undesirable. The shield problem may influence the scoring accuracy
Iqbal <i>et al.</i> (2011) for the Indonesian Sign Language (ISL).	DTW	The results show average accuracy of 95.60%	The selected words are only implied by one hand, i.e. the right hand
Asriani and Susilawati (2010) for the Indonesian Sign Language (ISL).	Backpropagation Neural Networks (BNN)	The success rate for static hand gesture recognition achieved in this study was 69.00%	The factor data cannot be recognized: the sign of static words nearly the same, the image of the faces more than dominant from the hand
Admasu and Raimond (2010) for Ethiopian Sign Language (ESL)	ANN	The experimental results achieve a recognition rate of 98.53%.	Complicated process and is not possible in real time
Hninn and Maung (2009): for Myanmar Sign Language	ANN	The system can achieve an average recognition rate of 90.00%.	The system is slower than the others but has fast execution time.
Stergiopoulou and Papamarkos (2009) Shape fitting technique	ANN	This system can achieve a recognition rate of 90.45% and 1.5 sec.	Two fingers may be classified into the same class of the finger. This problem has been overcome by choosing the most likely
Maraqqa and Abu-Zaiter (2008) for Arabic Sign Language	ANN	The system has had a performance with an accuracy rate of 95.00% for static gesture recognition	combinations of fingers. The feature extraction stage and the difficulty of determining the center of the hand for fingertip noise or where the image has been covered by a color.
Akmeiliawati <i>et al.</i> (2007) for Malaysian Sign Language system.	ANN	The system achieved a recognition rate of over 90.00%.	The signs that are similar in posture and gesture to another sign can be misinterpreted, resulting in a decrease in accuracy of the system.
Bailador <i>et al.</i> (2007) for real time gesture recognition	ANN	The system is fast, simple and modular. The recognition rate achieved 94.00% from testing the dataset.	The system cannot be used to predict all movements.
Binh and Ejima (2005) for hand gestures recognition.	fuzzy ARTMAP-NN	Experiments show that the system can achieved 92.19% recognition average rate.	Need to be developed to evaluate the proposed models on an online database of hand gesture images.
Bowden <i>et al.</i> (2004) for British Sign Language system (BSL)	HMM	Classification rates as high as 97.67% achieved for a lexicon of 43 words using only single instance training.	The system cannot recognize BSL sentences.
Al-Jarrah and Halawani (2001) for Arabic Sign Language.	ANFIS	The success rate of recognition accuracy was 93.55%.	using images of the bare hands.
Sekar (2001) for the Indonesian Sign Language (ISL) recognition	ANN	Results: 83.18% for 22 words (only using fingers + wrist) and 49.58% for 72 words (using all sensors).	The dataset comprises 72 cue words SIBI static. Used sensor flex.

## 5. CONCLUSION

In this study, we have presented a review of sign language recognition involving sign capturing and sign recognition methods. The literature knowledge gathered in this study is used to guide in choosing the best method for developing the Indonesian sign language recognition system.

Image-based and hand glove-based methods associated with a hand tracker are the methods most used for capturing hand gestures in sign language. Using data-gloves such as Kinect enables more intricate gestures that involve moving individual fingers, wrist and hands, allowing for more flexible, accurate and reliable gesture recognition. Meanwhile, the image-based method offers user-independent feature extraction, but needs more processing for feature extraction and noise reduction.



The major issues relate to the sign language translator system are accuracy and efficiency. Therefore, it is vital to use the right sign capturing method for integrating with the right sign language recognition methods. In addition, to produce a good sign language translator system, some researchers have used a combination sign language methods (e.g., HMM and DTW, Neural and Fuzzy).

To overcome the accuracy and efficiency problems in a sign language system, this study proposed the following solution: a sign language recognition method using hybrid Fuzzy and Neural Network with sign capturing based on Kinect camera.

## 6. ACKNOWLEDGMENT

This study was supported under the research grant No Vote GRS130339, Universiti Malaysia Pahang, Malaysia.

## 7. REFERENCES

- Admasu, Y.F. and K. Raimond, 2010. Ethiopian sign language recognition using artificial neural network. Proceedings of the 10th International Conference on Intelligent Systems Design and Applications, Nov. 29-Dec. 1, IEEE Xplore Press, Cairo, pp: 995-1000. DOI: 10.1109/ISDA.2010.5687057
- Akmeiliawati, R., M.P. Ooi and Y.C. Kuang, 2007. Real-time malaysian sign language translation using colour segmentation and neural network. Proceedings of the IEEE Instrumentation and Measurement Technology Conference Proceedings, May 1-3, IEEE Xplore Press, Warsaw, pp: 1-6. DOI: 10.1109/IMTC.2007.379311
- Al-Jarrah, O. and A. Halawani. 2001. Recognition of gestures in Arabic sign language using neuro-fuzzy systems. *Artif. Intell.*, 133: 117-138. DOI: 10.1016/s0004-3702(01)00141-2
- Asriani, F. and H. Susilawati, 2010. Static hand gesture recognition of Indonesian sign language system based on backpropagation neural networks. *MAKARA J. Technol. Series*, 14: 150-154.
- Bailador, G., D. Roggen and G. Tröster, 2007. Real time gesture recognition using continuous time recurrent neural networks. Proceedings of the ICST 2nd International Conference on Body Area Networks, Jun. 11-13, ICST, Brussels, Belgium, pp: 1-8. [dl.acm.org/citation.cfm?id=1460247](http://dl.acm.org/citation.cfm?id=1460247)
- Bellman, R. and R. Kalaba, 1959. On adaptive control processes. *IRE Trans. Automat. Control*, 4: 1-9. DOI: 10.1109/TAC.1959.1104847
- Binh, N.D. and T. Ejima, 2005. Hand gesture recognition using fuzzy neural network. Proceedings of the GVIP Conference, Dec. 19-21, GVIP Cairo, Egypt, pp: 7.
- Bowden, R., D. Windridge, T. Kadir, A. Zisserman and M. Brady, 2004. A Linguistic Feature Vector for the Visual Interpretation of Sign Language. In: *Computer Vision-ECCV 2004*, Pajdla, T. and J. Matas (Eds.), Springer Berlin Heidelberg, pp: 390-401.
- Braem, P.B., 1995. Einführung in die Gebärdensprache und ihre Erforschung. 2nd Edn., Signum Verlags GmbH, Hamburg, ISBN-10: 3927731102, pp: 232.
- Brashear, H., T. Starner, P. Lukowicz and H. Junker, 2003. Using multiple sensors for mobile sign language recognition. Proceedings of the 7th IEEE International Symposium on Wearable Computers, Oct. 18-21, IEEE Xplore Press, pp: 45-52. DOI: 10.1109/ISWC.2003.1241392
- Capilla, D.M., 2012. Sign language translator using microsoft kinect XBOX 360TM. Department of Electrical Engineering and Computer Science, University of Tennessee.
- DPN, 2001. Kamus Sistem Isyarat Bahasa Indonesia. 3rd Edn., Departemen Pendidikan Nasional, Jakarta, pp: 835.
- Dymarski, P., 2011. Hidden Markov Models, Theory and Applications. InTech.
- ElenaSanchez-Nielsen, L. Anton-Canalís and M. Hernandez-Tejera, 2003. Hand gesture recognition for human-machine interaction. *J. WSCG*, 12: 1-3
- Hasan, M.M. and P.K. Mishra, 2010. HSV brightness factor matching for gesture recognition system. *Int. J. Image Process.*, 4: 456-467.
- Hninn, T. and H. Maung, 2009. Real-time hand tracking and gesture recognition system using neural networks. *World Acad. Sci. Eng. Technol.*, 50: 466-470.
- Iqbal, M., E. Purnama and M.H. Purnomo, 2011. Indonesian sign language recognition system based on flex sensors and accelerometer using Dynamic Time Warping. *Teknik Elektro, Institut Teknologi Sepuluh Nopember*.
- Karami, A., B. Zanj and A. K. Sarkaleh, 2011. Persian Sign Language (PSL) recognition using wavelet transform and neural networks. *Expert Syst. Applic.*, 38: 2661-2667. DOI: 10.1016/j.eswa.2010.08.056.

- Lang, S., B. Marco and R. Raúl, 2011. Sign Language Recognition Using Kinect. In: L.K. Rutkowski, Marcin and R. T. Scherer, Ryszard Zadeh, Lotfi Zurada, Jacek (Eds.), Springer Berlin / Heidelberg, pp: 394-402.
- Li, H. and M. Greenspan, 2007. Continuous time-varying gesture segmentation by dynamic time warping of compound gesture models. Queen's University, Kingston, Canada.
- Ma, J., W. Gao, J. Wu and C. Wang, 2000. A continuous Chinese sign language recognition system. Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, Mar. 28-30, IEEE Xplore Press, Grenoble, pp: 428-433. DOI: 10.1109/AFGR.2000.840670
- Maraqqa, M. and R. Abu-Zaiter, 2008. Recognition of Arabic Sign Language (ArSL) using recurrent neural networks. Proceedings of the 1st International Conference on the Applications of Digital Information and Web Technologies, Aug. 4-6, IEEE Xplore Press, Ostrava, pp: 478-448. DOI: 10.1109/ICADIWT.2008.466439
- Mitra, S. and T. Acharya, 2007. Gesture recognition: A survey. IEEE Trans. Syst. Man Cybernet., 37: 311-324. DOI: 10.1109/TSMCC.2007.893280
- Ng, W., C. Ranganath and Surendra, 2002. Real-time gesture recognition system and application. Image Vision Comput., 20: 993-1007. DOI: 10.1016/S0262-8856(02)00113-0
- Sandjaja, I.N. and N. Marcos, 2009. Sign language number recognition. Proceedings of the 5th International Joint Conference on INC, IMS and IDC, Aug. 25-27, IEEE Xplore Press, Seoul, pp: 1503-1508. DOI: 10.1109/NCM.2009.357
- Sekar, E.T., 2001. perancangan dan implementasi prototipe sistem pengenalan bahasa isyarat. Magister ITB, Institut Teknologi Bandung.
- Septiari, R. and H. Haryanto, 2012. Converter sound with indonesian language as input to video sign language gesture using speech recognition method Hidden Markov Model (HMM) for Deaf People.
- Stergiopoulou, E. and N. Papamarkos, 2009. Hand gesture recognition using a neural network shape fitting technique. Eng. Applic. Artif. Intell., 22: 1141-1158. DOI: 10.1016/j.engappai.2009.03.008
- Xiaoyu, W., J. Feng and Y. Hongxun, 2008. DTW/ISODATA algorithm and multilayer architecture in sign language recognition with large vocabulary. Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Aug. 15-17, IEEE Xplore Press, Harbin, pp: 1399-1402. DOI: 10.1109/IIH-MSP.2008.136
- Xingyan, L., 2003. Gesture recognition based on fuzzy c-means clustering algorithm. Department of Computer Science the University of Tennessee Knoxville.
- Yanuardi, A.W., S. Prasetyo and P.P. Johannes Adi, 2010. Indonesian sign language computer application for the deaf. Proceedings of the 2nd International Conference on Education Technology and Computer, Jun. 22-24, IEEE Xplore Press, Shanghai, pp: V2-89-V82-92. DOI: 10.1109/ICETC.2010.5529427
- Zhang, L., J.C. Hsieh and J. Wang, 2012. A kinect-based golf swing classification system using HMM and neuro-fuzzy. Proceedings of the International Conference on Computer Science and Information Processing, Aug. 24-26, IEEE Xplore Press, Xi'an, Shaanxi, pp: 1163-1166. DOI: 10.1109/CSIP.2012.6309065