Pertanika J. Sci. & Technol. 26 (4): 1659 - 1675 (2018)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

A Survey of Hand Gesture Recognition Methods in Sign Language Recognition

Suharjito[1]*, Meita Chandra Ariesta[2], Fanny Wiryana[2] and Gede Putra Kusuma[1]

[1]*Computer Science Department, Bina Nusantara University (BINUS), Jakarta 11480, Indonesia,*
[2]*Computer Science Department, School of Computer Science, Bina Nusantara University (BINUS), Jakarta 11480, Indonesia*

ABSTRACT

Sign Language is the only method used in communication between the hearing-impaired community and common community. Sign Language Recognition (SLR) system, which is required to recognize sign languages, has been widely studied for years. The studies are based on various input sensors, gesture segmentation, extraction of features and classification methods. This paper aims to analyze and compare the methods employed in the SLR systems, classifications methods that have been used, and suggests the most promising method for future research. Due to recent advancement in classification methods, many of the recent proposed works mainly contribute on the classification methods, such as hybrid method and Deep Learning. This paper focuses on the classification methods used in prior Sign Language Recognition system. Based on our review, HMM-based approaches have been explored extensively in prior research, including its modifications. Deep Learning such as Convolutional Neural Network is popular in the past five years. Hybrid CNN-HMM and fully Deep Learning approaches have shown promising results and offer opportunities for further exploration. However, overfitting and high computational requirements still hinder their adoption. We believe the future direction of the research is toward developing a simpler network that can achieve high performance and requires low computational load, which embeds the feature learner into the classifier in multi-layered neural network fashion.

Keywords: Classification methods, Hidden Markov model, neural networks, sign language recognition

INTRODUCTION

Sign language is delivered through visual communications such as gestures, hand-shapes, facial expressions, and movement

ARTICLE INFO

Article history:
Received: 15 September 2017
Accepted: 17 April 2018
Published: 24 October 2018

E-mail addresses:
harjito@yahoo.com (Suharjito)
keroaduh@gmail.com (Meita Chandra Ariesta)
fannywirya@gmail.com (Fanny Wiryana)
inegara@binus.edu (Gede Putra Kusuma)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

of hands. Different from oral languages, sign language has its own vocabulary comprising combinations of various visual features for conveying messages in words or sentences (Sahoo, Mishra, & Ravulakollu, 2014). The communication between hearing impaired community and common community is limited mostly because of the common community's lack of knowledge about sign language. Thus, it is very hard for members of the common community to freely converse with hearing-impaired persons. This topic is very important as we are experiencing globalization and every person should receive the same opportunities regardless of their backgrounds and physical conditions. Therefore, a system which can translate sign language into common language and vice versa is needed. The system may help normal persons understand sign language so that the communication between hearing impaired and common communities can be easier.

This paper aims to analyze and compare the methods implemented in previous researches. Moreover, it aims to suggest the best method to explore for future research. We hope to create an Indonesian Sign Language Recognition system using the method suggested by this study. Several studies have been reported that review prior works in order to suggest the best method in Sign Language Recognition system. Majid and Zain (2013) reviewed the development of Sign Language Recognition system for different sign languages. They reviewed only 32 related publications up to year 2012. The review was focused on data acquisition and recognition methods. They suggested a sign language recognition system using hybrid Fuzzy and Neural Network with Kinect to tackle accuracy and efficiency problems. A comprehensive review on hand gesture recognition was reported by Pishardy et al. (Pisharady & Saerbeck, 2015). They reviewed 159 publications up to year 2015. The review paper contained prior works related to data acquisition, feature extraction, and classification methods. They also reviewed some hand gesture databases. They suggested that Time Delay Neural Network was compact and efficient to use because it optimized features detection and reduced training time, while Hidden Markov Model requires a large number of training data and high computational costs (Pisharady & Saerbeck, 2015). However, none the classification methods reviewed in their paper is about deep learning approach, which is recently getting popularity as a powerful classification method. In this work, we add some latest developments in the field including deep learning approaches, and provide more coverage on the classification methods. We reviewed 70 publications, among which 19 publications were from the year 2016 and 2017. These publications were obtained through Google Scholar search engine. We notice that many of the recent proposed works mainly contribute on the classification methods, such as hybrid method and deep learning, rather than on stages prior to the classification. Therefore, we focus our review on the classification methods. For completeness, we also report an overview of the pipeline of hand gesture recognition that consists of data acquisition, hand segmentation, feature extraction, and recognition methods in the next section.

Feature extraction, recognition methods, and implementation are essentials in the development of Sign Language Recognition (SLR) system. According to Bhuyan, Kumar, MacDorman and Iwahori (2014), determining the start and end point of a meaningful gesture, also called gesture segmentation, was one of many challenges in hand gesture recognition. Feature extraction is another problem. Zhang et al. optimized feature extraction process by detecting pupil and using it as a reference point as well as using colored-gloves for hand detection (Zhang, Chen, Fang, Chen, & Gao, 2004). In recent studies, Kinect had been widely used for image acquisitions in SLR system because it could track hand and body actions easily and accurately. It also provided depth and color data at the same time (Chai et al., 2013). On top of that, research on the classification method was also elaborated. Although feature extraction is important, the classification method employed is also important and still needs improvements. To overcome the difficulties in hand segmentation, hand tracking, and complex backgrounds in SLR systems, Huang, Zhou, Li and Li (2015) et al. implemented 3D Convolutional Neural Network to address these problems. There are many studies that experimented on various classification methods in SLR system. The methods used in the previous studies will be analyzed and compared further in this paper to find the best classification method, which will be used in an Indonesian Sign Language Recognition system in the future.

Hand Gesture Recognition

Ahuja and Singh (2015) explained the basic module of Hand Gesture Recognition. The basic module of Hand Gesture Recognition consists of four steps: Image Acquisition, Hand Segmentation, Feature Extraction, and Hand Gesture Recognition, as shown in Figure 1. Image acquisition is the process of capturing images for vision-based approach. Leap Motion Controller, Microsoft Kinect, Data Glove, and Vision-based are some other methods of sign acquisition. Leap Motion Controller (LMC) is a device that can detect hand movement up to 200 frames per second and assign ID to each detection (Mohandes, Aliyu, & Deriche, 2014). LMC converts signals into computer commands (Bhavsar, 2017; Kakde, Nakrani, & Rawate, 2016). LMC has been commonly used in prior studies related to hand gesture recognition (Koul, Patil, Nandurkar, & Patil, 2016; Potter, Araullo, & Carter, 2013). Microsoft Kinect is used to capture every motion and turns it into a feature by using the built-in 3D sensory camera. Microsoft Kinect is widely used in hand gesture recognition (Carneiro et al., 2016; Escobedo & Camara, 2016; Jiang, Tao, Ye, Wang, & Ye, 2014; Keskin, Kırac, Kara, & Akarun, 2011; Raheja, Mishra, & Chaudhary, 2016; Rakun, Andriani, Wiprayoga, Danniswara, & Tjandra, 2013). Data-Glove Based Method is a relatively old data acquisition method for gesture recognition. This method utilizes a device to help the process of collecting data. The device is a glove which has sensors connected to a computer. Those sensors detect the movement and changes in the hands and fingers of the users (Mehdi, 2002; Phi, Nguyen, Bui, & Vu, 2015; Ranjini & M, 2014; Saengsri, Niennattrakul, & Ratanamahatana, 2012).

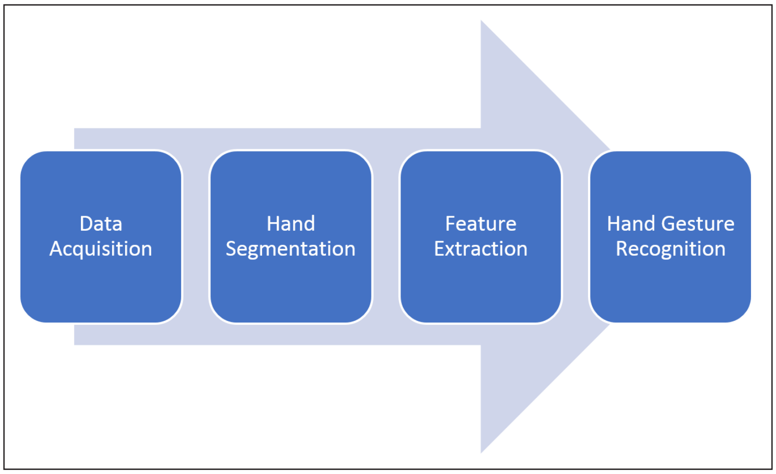

Figure 1. Basic hand gesture recognition module

Hand segmentation is a process to isolate hands and other features from the rest of the image in vision based systems. Zhang et al. (2004) employed pupil-detection algorithm to make the pupil as reference points and used colored gloves to assist the background segmentation. Many hand segmentation methods have been proposed in computer vision. Canny Edge Detector is used to detect the edges of hands from an image. Canny Edge Detector is known for its optimal performance in detecting edges and low error rate (Ghosh & Ari, 2016; Kalsh & Garewal, 2013; M. V. D. Prasad, Kishore, Kumar, & Kumar, 2016). The other method for hand segmentation is Elliptical Fourier Descriptors which is specialized for extracting outline of shapes (Kishore, Prasad, Prasad, & Rahul, 2015). Skin detection is also used in hand segmentation. This method simply identifies and segments the skin area from the rest of the image (Lim, Tan, & Tan, 2016; Ong & Ranganath, 2005). Pankajakshan & Thilagavathi (2015) also applied hand motion tracking with skin detection to produce more precise result Similar to skin detection, colored-gloves were used to give the hands a distictive feature, therefore assisting the hand segmentation process (Eugene Starner, 1995; Wang & Popović, 2009; Zhang, Chen, Fang, Chen, & Gao, 2004).

Feature extraction is used to acquire features from the images captured. The features include background data, translation, scale, shape, rotation, angle, coordinates, and movements (Bhavsar & Trivedi, n.d.). Yan et al. (2017) proposed multiscale Convolutional Neural Network to detect hands to tackle problems in prior studies regarding hand detection, which struggled due to low resolution, varying lighting intensity, heavy and occlusion. Scale

Invariant Feature Transform (SIFT) algorithm had been used as feature extraction methods in prior studies (Goyal & Singh, 2014; S. Goyal, Sharma, & Sharma, 2013). According to Gurjal and Kunmur (2012), SIFT algorithm was robust against rotation, translation, or scaling variation, and produced a large collection of local feature vectors.

Classification is the final stage and an important stage in gesture recognition (Ghosh & Ari, 2016). Word or sentence in sign language is made up of continuous gesture, which changes over time. Therefore, a recognition method should be able to handle sequential data. Some problems occur when the system handles noisy data and uncontrolled environment. Therefore, Roh et al. (2016) proposed Selective Temporal Filtering to tackle those problems. There are two types of Gesture Recognition approaches. Some researchers used the extracted features for gesture recognition such as template matching (Hernandez-Rebollar, Kyriakopoulos, & Lindeman, 2004; Keni, Meher, & Marathe, 2013) and some used machine learning classifiers such as Hidden Markov Model (HMM) which would be further discussed in the next section. According to Khan and Ibraheem (2012), Hand Gesture Recognition could be applied in different applications, including sign language recognition system, robot control, graphic editor, virtual environments (VE), numbers recognition, television control, and 3D modelling. These applications implemented hand gesture recognition system methods to perform the basic function of the application. One example of a sign language recognition system utilized hand gesture recognition module. After the image/video is recorded, the stream of data is loaded and segmented. Then, feature extraction is conducted in regards of shape, size, hand and finger trajectory, and body of the signer. HMM requires feature vectors; therefore, feature vector is created by using this information. Then, classification is performed by Hidden Markov Model. Feature vector must be extracted from each video frame for both training and recognition, then inputted into the Hidden Markov Model (Yang, Tao, Xi, & Ye, 2015).

METHODS

Classification Methods and Related Works

Hidden Markov Model (HMM). Several previous studies implement Hidden Markov Model (HMM) as the basis of research to make Sign Language Recognition (SLR) system. HMM is a statistic model where a set of parameters is hidden. The hidden parameters can be acquired from related observation parameters (Zabulis, Baltzakis, & Argyros, 2009). "Hidden Markov Model is a finite model that describes a probability distribution over an infinite number of sequences" (Eddy, 1996). HMM has been widely employed in Speech Recognition system. HMM is also used in glove-based Sign Language Recognition system (Liang & Ouhyoung, 1998; Ma, Gao, Wu, & Wang, 2000; H. Wang, Leu, & Oz, 2006). Hidden Markov Model is used to tackle sequential data and Sign Language consists of continuous gestures which make up a word or sentence.

Multi-dimensional Hidden Markov Model is used in recognizing American Sign Language (ASL) in, which has 96.7% accuracy (Wang et al., 2006). The data from input devices (CyberGlove™) is in the form of 21 data-stream, which is then segmented into gestures in the same interval. Subsequently, the data is inputted into 21-dimensional feature vector. Finally, the data is classified and recognized according to stochastic data to produce the output (recognized sign language). Similar previous research (Ma et al., 2000) also used HMM to recognize Chinese Sign Language (CSL) and was able to get 98.2% accuracy with embedded training. In an earlier research, Liang and Ouhyoung (1998) used HMM to recognize data from DataGlove™. The system could recognize 250 words of Taiwanese Sign Language (TSL) formed by 51 basic postures, 6 orientations, and 8 fundamental motions with 80.4% accuracy of real-time continuous gestures. The input was statistically analyzed by 4 parameters. The parameters are position, posture, motion, and orientation.

Starner and Pentland (1997) also created a real-time SLR system. However, their research used video-based approach. The research conducted two experiments by using 40 words lexicon, where in the first experiment the user wore coloured-glove and in the second experiment without coloured-glove. The first experiment reached 99% of accuracy, whilst the second experiment reached 92% of accuracy. Elmezain, Al-Hamadi, Appenrodt, & Michaelis (2008) created a real-time SLR system for 10 Arabic numbers in Sign Language using HMM from coloured image sequence and attained 98.94% of accuracy. There are SLR systems using HMM from various countries. The previous research includes Sign Language Recognition system for Taiwanese Sign Language (Lee, Yeh, & Hsiao, 2016; Liang & Ouhyoung, 1998), Chinese Sign Language (Ma et al., 2000), American Sign Language (Starner & Pentland, 1997; Wang et al., 2006), and Arabic Sign Language (Elmezain et al., 2008). Moreover, there is also Greek Sign Alphabet Letters Recognition system (Pashaloudi & Margaritis, 2004), which is able to recognize 90.20% of training set and 86.52% of testing set. The sets include 16 Greek Sign Alphabet Letters and the system recognizes pictures. Hidden Markov Model struggles handling noisy data (Roh et al., 2016); therefore, Kaluri & Pradeep, (2017) proposed the implementation of Wiener Filter to eliminate the noise in an images and Adaptive Histogram technique to segment the images which would be feed into HMM for training and recognition.

Modifications of Hidden Markov Model. Modification of Hidden Markov Model (HMM) has been researched to better improve the performance and accuracy of a Sign Language Recognition (SLR) system. For instance, a 3 state left-to-right Hidden Markov Model with three independent Gaussian Mixtures (GMM) and a globally merged covariance matrix is implemented in research with 17% error rate (Dreuw, Rybach, Deselaers, Zahedi, & Ney, 2007). This research implements speech recognition techniques to create an automatic Sign

Language Recognition system which can adapt to dialects in the sign language. Principal Component Analysis (PCA) is used to lessen the dimension and help the classification process of features obtained. "Principal Component Analysis is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set" (Ringnér, 2008). PCA is commonly used in hand recognition systems (Li, Kao, & Kuo, 2016; Prasad, Kishore, Kumar, & Kumar, 2016; Sawant, 2014; Zaki & Shaheen, 2011). PCA is also used to characterize the feature of fingers in a vision-based Chinese SLR system (Zhang et al., 2004). Tied-Mixture Density Hidden Markov Model (TMDHMM) is used to speed up the recognition system without significantly reducing the accuracy. The system could recognize up to 92.5% of the frequently used Chinese Sign Language. TMDHMM is also used because of the efficient computational costs. In (Yang et al., 2015), Weighted Hidden Markov Model assigns weights for each sign samples. This system used Kinect to improve recognition rate with 156 isolated sign words used as data. It attained a high recognition rate of up to 94.74%.

The latest studies have several Hidden Markov Model based improvements on the previous research. The research includes a hybrid CNN-HMM (Koller, Zargaran, Ney, & Bowden, 2016), Coupled-HMM (CHMM) (Kumar, Gauba, Roy, & Dogra, 2017), and Kinect-Based using Hidden Markov Model (Lee et al., 2016) for SLR system. CHMM was used in an SLR system and was able to reach 90.80% of accuracy (Kumar et al., 2017). This research used Kinect to create 3D model of the captured gestures. Another research also used Kinect for HMM based Taiwanese Sign Language Recognition (SLR) system and was able to attain 85.14% of recognition rate (Lee et al., 2016). HMM was used to determine the signing direction, whilst for the recognition, a trained SVM was used.

Table 1 summarizes the SLR using HMM and modified HMM. As shown in the table, the highest reported accuracy was achieved by Elmezain et al. (2008) on a small size dataset. The system could recognize 0-9 Arabic numbers in Arabic Sign Language. Some of the system could only recognize alphabets, numbers, and basic hand shapes (Elmezain et al., 2008; Pashaloudi & Margaritis, 2004; Wang et al., 2006) with high accuracy. There are also Sign Language Recognition system which can recognize a large number of vocabularies (Dreuw et al., 2007; Liang & Ouhyoung, 1998; Ma et al., 2000; Zhang et al., 2004). The system could recognize 220 words and 80 sentences in Chinese Sign Language with 98.2% of accuracy (Ma et al., 2000). However, one of the best systems from all of the systems mentioned before could recognize 92.5% of 439 words from Chinese Sign Language (Zhang et al., 2004). The sample used was 1756 words for training and 439 for testing. Weighted Hidden Markov Model is also worth mentioning as it outperformed other methods as well as the traditional Hidden Markov Model (Yang et al., 2015).

Table 1
Summary of SLR using HMM and modified HMMs

Author	Method(s)	Data	Result (%)
(Wang et al., 2006)	Multi-dimensional HMM	26 ASL alphabets and 36 ASL handshapes	96.7
(Ma et al., 2000)	HMM	220 CSL words and 80 CSL sentences	98.2
(Liang & Ouhyoung, 1998)	HMM	250 TSL words	80.4
(Starner & Pentland, 1997)	HMM	40 ASL words lexicon	92 – 99
(Elmezain et al., 2008)	**HMM**	10 numbers in Arabic Sign Language	**98.94**
(Pashaloudi & Margaritis, 2004)	HMM	16 letters in GSL	86.52 - 90.20
(Dreuw et al., 2007)	3 states left-to-right HMM	RWTH-Boston-104 corpus (201 sequence, 104 words)	17 (error rate)
(Zhang et al., 2004)	TMDHMM	439 signs in CSL	92.5
(Koller et al., 2016)	CNN-HMM	-	-
(Kumar et al., 2017)	CHMM	25 dynamic words from Indian Sign Language	90.80
(Lee et al., 2016)	HMM-SVM	12 directions from TSL	85.14
(Yang et al., 2015)	WHMM	156 isolated signs in CSL	94.74

Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is parallel computational models that simulate human brain, where processing nodes are called "neurons". Every neurons stored information and depend on it, it receives input from and send output to other neurons (Dongare, Kharde, & Kachare, 2012). (Adithya, Vinod, & Gopalakrishnan, (2013) employed ANN forward-backward algorithm to automatically recognize 26 alphabets and 10 numbers of Indian Sign Language with 91.1% of accuracy. However, the system only worked statically, not in real-time situation. Backpropagation Neural Network is the family of neural network models, where the learning algorithm is based on Deepest-Descent technique (Buscema, 1998). It was employed to build Sign Language Recognition (SLR) system for Indian Sign Language and attained 92.34% of recognition rate (Prasad, Kishore, & Kumar, 2016). The data used are own-made 80 sequences of video which in total consist of 59 signs of letters, numbers, and 23 words in Indian sign language.

Convolutional Neural Network (CNN)

"Convolutional Neural Network is the family of neural network models that feature a type of layer known as the convolutional layer which can extract features" (Pham, Kruszewski, & Boleda, 2016). CNN has been widely used in Computer Vision projects. Several of those projects used CNN to recognize Sign Language. Huang et al. (2015) used 3D CNN to recognize Sign Language into texts or speech. The accuracy of the system was as high as 94.2%. The data used consisted of 25 words of sign language used in daily conversations. Pigou, Dieleman, Kindermans, & Schrauwen, (2014) used CNN for Italian SLR. The system was automated due to Convolutional Neural Network and used Microsoft Kinect and GPU acceleration. The system could recognize 20 Italian Gestures with the accuracy of 91.7%. A Hybrid CNN-Hidden Markov Model (CNN-HMM) was used for continuous SLR system (Koller et al., 2016). Hidden Markov Model has the capabilities of sequence modelling. It can be combined with CNN, which has discriminative strength. The best system using CNN had 94.2% of recognition rate; however, the data used was small (Huang et al., 2015). 3D Convolutional Neural Network is employed in the creation of LipNet, which is a deep neural network that could do lipreading through visual approach. For the first time, LipNet could recognize end-to-end sentence-level lipreading and achieved state-of-the-art result of 95.5%. LipNet architecture consisted of 3D convolutional Neural Network, Bidirectional RNN, SOFTMAX, and Connectionist Temporal Classification (CTC). The experiments were performed on GRID corpus in which 28775 videos were used for training and 3971 videos were used for testing (Assael, Shillingford, Whiteson, & de Freitas, 2016).

Self-organizing Map (SOM)

"The Self-Organizing Map (SOM) is a software tool for the visualization of high-dimensional data. It implements an orderly mapping of a high-dimensional distribution onto a regular low-dimensional grid" as defined by (Kohonen, 1998). A Kohonen Self-Organizing Map was used in an Indian Sign Language Recognition system. SOM is sometimes also called Kohonen Self-Organizing Feature Map (SOFM). Neural Network is used for pattern recognition run in MATLAB. The system attained a maximum accuracy of 80% from 35 images of 5 sign language captured in 7 different backgrounds, with 72 seconds of training time. Gao et al. (2004) employed Self-Organizing Map together with other methods such as Hidden Markov Model (HMM) and Simple Recurrent Network (SRN). Simple Recurrent Network was implemented to segment continuous sign language according to Self-Organizing Feature Map (SOFM) representations. The output of Simple Recurrent Network (SRN) was then used as HMM states. Viterbi Algorithm is employed to search the best matched word of the according signs. The highest accuracy was 91.3% in embedded training and registered test sets.

Other Methods

Korean Sign Language Recognition system was able to attain 85% of recognition rate out of 25 Korean Sign Language (KSL) word (Kim, Jang, & Bien, 1996). The system employed Fuzzy Min-Max to recognize the input from a Data-glove. The data received was the position of hand (x, y, z axes) forming 10 primary movements and 14 basic hand shapes. "Fuzzy Min-Max Classification Neural Network is Hyper boxes defined by pairs of min-max points, and their corresponding membership functions are used to create fuzzy subsets of the n-dimensional pattern space" as defined by (Simpson, 1992). Template Matching is often used in Computer Vision, also used in Sign Language Recognition system (Hernandez-Rebollar et al., 2004; Keni et al., 2013). Helped by 17 volunteers with different level of skills to demonstrate 30 American Sign Language words, the system in (Hernandez-Rebollar et al., (2004) attained 98% of accuracy. Ghosh and Ari (2016) proposed an enhanced version of Radial Basis Function (RBF) Neural Network for the classification of hand gestures. K-means algorithm was used to select the centers of RBF classifier automatically and least-mean-square (LSM) algorithm is utilized to update the estimated weight matrix recursively. K-Nearest Neighbors (KNN) classifier was used along with PCA, generating 96.31% highest accuracy. The system consisted of three phases: hand segmentation, feature extraction, and classification. To capture the image depth, the system used Kinect. The database used consisted of 61 hand gestures, each from 10 different signers totaling 12,200 images of Brazilian Sign Language (Costa Filho, Souza, Santos, Santos, & Costa, 2017). KNN classifier also produces a high result of 99.61% accuracy in recognizing Indian Sign Language. The data used consisted of 3600 images from 40 different signers. The data varies in lighting condition, angle, and distance. Lim, Tan and Tan (2016) handled the problems in sign language recognition using feature covariance matrix to recognize isolated sign language with 87.33% recognition rate for ASL. Feature covariance matrix is able to reduce the dimension of features and combine associated signs naturally. Roh et al. (2016) proposed a novel approach in handling noisy data by using Selective Temporal Filtering (STF). In a not noisy environment, the system reached 92.1% of accuracy; meanwhile, in a noisy environment, the accuracy dropped to 62.2%. Their study found that STF performs better than HMM and Conditional Random Fields not only in a noisy environment but also in a balanced environment.

RESULTS AND DISCUSSION

There are several notable prior works that represent the direction of the classification methods used in SLR system. Ma et al. (2000) proposed a HMM based sign language recognizer that could achieve high accuracy results in recognizing Chinese Sign Language sentences. Gesture inputs were obtained using two Cybergloves and two 3SAPCE-position trackers, and invariant features were extracted based on 3D movements of the signer. Using the combination of dynamic programing, HMM, Bigram language model, and search algorithm, they managed to recognize CSL at a sentence level at 98.2% accuracy. However, their SLR system still relied on active sensors for data acquisition to achieve high accuracy results. HMM-based SLR approaches have been shown to achieving good recognition accuracy especially in small to medium-sized datasets. Efforts on improving the performance of HMM-based approaches have also been proposed by modifying the standard HMM method. Zhang et al. proposed the tied-mixture density HMM (TMDHMM) to increase the recognition speed without sacrificing the recognition accuracy (Zhang et al., 2004). Meanwhile, Yang et al. (2015) proposed weighted HMM (WHMM) to cope with sign variations from different signers. Both modifications can achieve more than 90% accuracy on medium-sized datasets of word-level signs. Nevertheless, the proposed HMM-based and modified HMM-based approaches are still required to transform the input data into "handcrafted" sign features before applying the classification method. Designing invariant sign features can be tedious works and highly dependent on the type of input data being used. Moreover, feature extraction can also be seen as an additional step potentially slowing down the recognition speed and the classification performance often depends on the quality of the extracted sign features.

Koller et al. (2016) proposed a Hybrid CNN-HMM approach in order to combine the discriminative strength of the CNN and the sequential modeling of the HMM. In this hybrid approach, they used the output posteriors of the CNN as observation probabilities of the HMM. Therefore, it enabled them to perform end-to-end training. They also compared the performance of the Hybrid CNN-HMM to the Tandem CNN-HMM approach, where CNN was treated as a feature extractor for the HMM classifier. Their experimental results showed that the Hybrid CNN-HMM outperformed the Tandem CNN-HMM approach. The current implementation of the hybrid approach still focuses only on the right hand, which is assumed to be the dominant hand of the signer. This method also requires a good frame-state alignment. An attempt to break away from HMM-based approach was done by Huang et al. (2015). They proposed a 3D CNN to recognize sign language directly from input data without the need of designing handcrafted features. The input data was obtained from a Microsoft Kinect, which contained color and depth information, and the body joint positions of the signer. The proposed method had shown to achieve promising results. However, the results were obtained from a small-sized dataset. This approach also cannot handle variable-length sequences of frames. Additional pre-processing is necessary to select a fixed number of frames for each input channel. Recently, a promising deep learning approach for sentence-level sequence prediction has been proposed by Assael, Shillingford, Whiteson and de Freitas (2016). They combined the 3D CNN and Bidirectional RNN to perform lipreading at the sentence-level. The proposed approach was evaluated on large-sized dataset and achieved 95.5% accuracy. Lipreading task is very similar to sign language recognition where both try to interpret spatio-temporal cues obtained from input video to predict spoken words or sentences. A similar approach can be employed to solve SLR problem. However, it tends to overfit when trained on small dataset due to complexity of the deep model. Also, it requires high computational load. It is of great interest to develop a simpler network that can achieve high performance and requires low computational load, but still embeds the feature learner into the classifier in multi-layered neural network fashion.

CONCLUSION AND FUTURE WORK

Different approaches have been proposed by various studies to solve the problem of sign language recognition (SLR). We have reviewed prior works based on different stages of the recognition procedure, which includes data acquisition, gesture segmentation, feature extraction, and classification. In this contribution, we provide more coverage on the classification approaches. We have reviewed many prior works such as HMM-based, modified HMM based, neural network based, and hybrid-based approaches. HMM-based SLR approaches have been shown to achieving good recognition accuracy especially in small to medium sized datasets. Efforts on improving the performance of HMM-based approaches have also been proposed by modifying the standard HMM method. However, the proposed HMM-based and modified HMM-based approaches still require the extraction of sign features from input data before applying the classification method. Designing invariant sign features can be tedious works highly dependent on the type of input data being used. Moreover, feature extraction also contributes to the computational load and the classification performance often depends on the quality of the extracted sign features.

Combinations of CNN and HMM have also been proposed to improve the performance of the SLR system. It has been shown that a hybrid approach that embeds CNN into HMM abiding to Bayesian principles outperforms the tandem approach that treats CNN as a feature extractor. The hybrid approach also enables the end-to-end training. The 3D CNN and the combination of 3D CNN and RNN have recently been shown to potentially improve the SLR performance. However, overfitting and high computational requirements still hinder their adoption. We believe the future direction of the research is toward developing a simpler network that can achieve high performance and requires low computational load, which embeds the feature learner into the classifier in multi-layered neural network fashion.

ACKNOWLEDGEMENT

This research was supported by Research Grant No. 039A/VR.RTT/VI/2017 from Ministry of Research Technology and Higher Education of Republic of Indonesia.

REFERENCES

Adithya, V., Vinod, P. R., & Gopalakrishnan, U. (2013). Artificial neural network based method for Indian sign language recognition. In *2013 IEEE Conference on Information & Communication Technologies (ICT)* (pp. 1080–1085). Thuckalay, Tamil Nadu, India.

Ahuja, M. K., & Singh, A. (2015). A Survey of Hand Gesture Recognition. *International Journal*, *3*(5), 266-271.

Apoorva Ranjini, S. S., Chaitra, M., Deepika, V., & Patil, J. M. (2014). Sign Language Recognition System. *International Journal on Recent and Innovation Trends in Computing and Communication*, *2*(4), 947 – 953.

Assael, Y. M., Shillingford, B., Whiteson, S., & de Freitas, N. (2016). LipNet: Sentence-level Lipreading. Computing Research Repository (CoRR), 1611.01599, 1-13. Retrieved July 4, 2017 from http://arxiv.org/abs/1611.01599.

Bhavsar, H. (2017). Review on Classification Methods used in Image based Sign Language Recognition System. *International Journal on Recent and Innovation Trends in Computing and Communication*, *5*(5), 949–959.

Bhavsar, H., & Trivedi, J. (2017). Review on Feature Extraction methods of Image based Sign Language Recognition system. *Indian Journal of Computer Science and Engineering (IJCSE), 8*(3), 249-259.

Bhuyan, M. K., Kumar, D. A., MacDorman, K. F., & Iwahori, Y. (2014). A novel set of features for continuous hand gesture recognition. *Journal on Multimodal User Interfaces*, *8*(4), 333–343.

Buscema, M. (1998). Back propagation neural networks. *Substance Use & Misuse*, *33*(2), 233–270.

Carneiro, S. B., Ferreira, J. O., Barbosa, T. M., Da Rocha, A. F., Soares Alcala, S. G., & M. Santos, E. D. (2016). Static Gestures Recognition for Brazilian Sign Language with Kinect Sensor. In *2016 IEEE SENSORS* (pp. 1-3). Orlando, FL, USA.

Chai, X., Li, G., Lin, Y., Xu, Z., Tang, Y., Chen, X., & Zhou, M. (2013). Sign language recognition and translation with kinect. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 1-2). Shanghai, China.

Costa Filho, C. F. F., Souza, R. S. de, Santos, J. R. dos, Santos, B. L. dos, & Costa, M. G. F. (2017). A fully automatic method for recognizing hand configurations of Brazilian sign language. *Research on Biomedical Engineering*, *33*(1), 78–89.

Dongare, A. D., Kharde, R. R., & Kachare, A. D. (2012). Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology*, *2*(1), 189–194.

Dreuw, P., Rybach, D., Deselaers, T., Zahedi, M., & Ney, H. (2007). Speech recognition techniques for a sign language recognition system. In *Eighth Annual Conference of the International Speech Communication Association* (pp. 2513-2516). Antwerp, Belgium.

Eddy, S. R. (1996). Hidden markov models. *Current Opinion in Structural Biology*, *6*(3), 361–365.

Elmezain, M., Al-Hamadi, A., Appenrodt, J., & Michaelis, B. (2008). A hidden markov model-based continuous gesture recognition system for hand motion trajectory. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (pp. 1–4). Tampa, FL, USA.

Escobedo, E., & Camara, G. (2016). A new Approach for Dynamic Gesture Recognition using Skeleton Trajectory Representation and Histograms of Cumulative Magnitudes. In *SIBGRAPI Conference on Graphics, Patterns and Images* (pp. 209–216). São Paulo, Brazil.

Eugene Starner, T. (1995). *Visual Recognition of American Sign Language Using Hidden Markov Models*. Massachusetts Inst Of Tech Cambridge Dept Of Brain And Cognitive Science.

Gao, W., Fang, G., Zhao, D., & Chen, Y. (2004). A Chinese sign language recognition system based on SOFM/SRN/HMM. *Pattern Recognition*, *37*(12), 2389–2402.

Ghosh, D. K., & Ari, S. (2016). On an algorithm for Vision-based hand gesture recognition. *Signal, Image and Video Processing*, *10*(4), 655–662.

Goyal, E. K., & Singh, A. (2014). Indian Sign Language Recognition System for Deaf People. *Journal on Today's Ideas - Tomorrow's Technologies*, *2*(2), 145–151.

Goyal, S., Sharma, I., & Sharma, S. (2013). Sign Language Recognition System for Deaf and Dumb People. *International Journal of Engineering Research and Technology*, *2*(4), 382–387.

Gurjal, P., & Kunnur, K. (2012). Real time hand gesture recognition using SIFT. *International Journal of Electronics and Electrical Engineering*, *2*(3), 19–33.

Hernandez-Rebollar, J. L., Kyriakopoulos, N., & Lindeman, R. W. (2004). A new instrumented approach for translating American Sign Language into sound and text. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 547–552). Seoul, South Korea.

Huang, J., Zhou, W., Li, H., & Li, W. (2015). Sign language recognition using 3D convolutional neural networks. In *Multimedia and Expo (ICME), 2015 IEEE International Conference on* (pp. 1–6). Turin, Italy.

Jiang, Y., Tao, J., Ye, W., Wang, W., & Ye, Z. (2014). An Isolated Sign Language Recognition System Using RGB-D Sensor with Sparse Coding. In *2014 IEEE 17th International Conference on Computational Science and Engineering* (pp. 21–26). Chengdu, China.

Kakde, M. U., Nakrani, M. G., & Rawate, A. M. (2016). A Review Paper on Sign Language Recognition System For Deaf And Dumb People using Image Processing. *International Journal of Engineering Research and Technology*, *5*(3), 590–592.

Kalsh, E. A., & Garewal, N. S. (2013). Sign Language Recognition System. *International Journal of Computational Engineering Research*, *3*(6), 15-21.

Kaluri, R., & Pradeep, C. H. (2017). An enhanced framework for sign gesture recognition using hidden Markov model and adaptive histogram technique. *International Journal of Intelligent Engineering & Systems, 10*(3), 11-19.

Keni, M., Meher, S., & Marathe, A. (2013). Sign Language Recognition System. *International Journal of Scientific and Engineering Research*, *4*(12), 580-583.

Keskin, C., Kırac, F., Kara, Y. E., & Akarun, L. (2011) Real Time Hand Pose Estimation using Depth not according to formarSensors. In *IEEE International Conference on Computer Vision Workshops* (pp. 1228–1234). Barcelona, Spain.

Khan, R. Z., & Ibraheem, N. A. (2012). Hand gesture recognition: a literature review. *International Journal of Artificial Intelligence & Applications*, *3*(4), 161.

Kim, J. S., Jang, W., & Bien, Z. (1996). A Dynamic Gesture Recognition System For The Korean Sign Language (KSL). *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *26*(2), 354–359.

Kishore, P. V. V, Prasad, M. V. D., Prasad, C. R., & Rahul, R. (2015). 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN. In *International Conference on Signal Processing and Communication Engineering Systems - Proceedings of SPACES 2015, in Association with IEEE* (pp. 34–38). Guntur, India.

Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, *21*(1), 1–6.

Koller, O., Zargaran, O., Ney, H., & Bowden, R. (2016). Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. In *Proceedings of the British Machine Vision Conference 2016*. New York, UK.

Koul, M., Patil, P., Nandurkar, V., & Patil, S. (2016). Sign Language Recognition Using Leap Motion Sensor. *International Research Journal of Engineering and Technology (IRJET)*, *3*(11), 322–325.

Kumar, P., Gauba, H., Roy, P. P., & Dogra, D. P. (2017). Coupled hmm-based multi-sensor data fusion for sign language recognition. *Pattern Recognition Letters*, *86*, 1–8.

Lee, G. C., Yeh, F.-H., & Hsiao, Y.-H. (2016). Kinect-based taiwanese sign-language recognition system. *Multimedia Tools and Applications*, *75*(1), 261–279.

Li, T. H. S., Kao, M. C., & Kuo, P. H. (2016). Recognition System for Home-Service-Related Sign Language Using Entropy-Based K-Means Algorithm and ABC-Based HMM. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *46*(1), 150–162.

Liang, R. H., & Ouhyoung, M. (1998). A real-time continuous gesture recognition system for sign language. In *Third IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 558–567). Nara, Japan.

Lim, K. M., Tan, A. W. C., & Tan, S. C. (2016). A feature covariance matrix with serial particle filter for isolated sign language recognition. *Expert Systems with Applications*, *54*, 208–218.

Ma, J., Gao, W., Wu, J., & Wang, C. (2000). A Continuous Chinese Sign Language Recognition System. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000* (pp. 428–433). Washington, DC, USA

Majid, M. A., & Zain, J. M. (2013). A Review on The Development of Indonesian Sign Language Recognition System. *Journal of Computer Science*, *9*(11), 1496–1505.

Mehdi Y. N., S. A. K. (2002). Sign language recognition using sensor gloves. In *Proceedings of the 9th International Conference on Neural Information* (pp. 2204–2206). Singapore.

Mohandes, M., Aliyu, S., & Deriche, M. (2014). Arabic sign language recognition using the leap motion controller. In *23rd International Symposium on Industrial Electronics (ISIE)* (pp. 960–965). Istanbul, Turkey.

Ong, S. C. W., & Ranganath, S. (2005). Automatic Sign Language Analysis : A Survey and the Future beyond Lexical Meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(6), 873–891.

Pankajakshan, P. C., & Thilagavathi, B. (2015). Sign Language Recognition System. In *International conference on Innovation in Information Embedded and Communication System* (pp. 2–5). Coimbatore, India.

Pashaloudi, V. N., & Margaritis, K. G. (2004). A performance study of a recognition system for greek sign language alphabet letters. In *9th Conference Speech and Computer* (pp. 545-551). Saint-Petersburg, Russia.

Pham, N. Q., Kruszewski, G., & Boleda, G. (2016). Convolutional Neural Network Language Models. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1153–1162). Austin, Texas.

Phi, L. T., Nguyen, H. D., Bui, T. T. Q., & Vu, T. T. (2015). A glove-based gesture recognition system for Vietnamese sign language. *Automation and Systems, Proceedings*, *13*(16), 1555–1559.

Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2014). Sign language recognition using convolutional neural networks. In *Workshop at the European Conference on Computer Vision* (pp. 572–578). Cham : Springer.

Pisharady, P. K., & Saerbeck, M. (2015). Recent methods and databases in vision-based hand gesture recognition: A review. *Computer Vision and Image Understanding*, *141*, 152–165.

Potter, L. E., Araullo, J., & Carter, L. (2013). The Leap Motion controller. In *Proceedings of the 25th Australian Computer-Human Interaction Conference on Augmentation, Application, Innovation, Collaboration* (pp. 175–178). Adelaide, Australia.

Prasad, D., Kishore, V., & Kumar, A. (2016). Indian sign language recognition system using new fusion based edge operator. *Journal of Theoretical and Applied Information Technology*, *88*(3), 574–584.

Raheja, J. L., Mishra, A., & Chaudhary, A. (2016). Indian Sign Language Recognition Using SVM 1. *Pattern Recognition and Image Analysis*, *26*(2), 434–441.

Rakun, E., Andriani, M., Wiprayoga, I. W., Danniswara, K., & Tjandra, A. (2013). Combining depth image and skeleton data from kinect for recognizing words in the sign system for indonesian language (sibi [sistem isyarat bahasa indonesia]). In *2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 387–392). Bali, Indonesia.

Ringnér, M. (2008). What is principal component analysis? *Nature Biotechnology*, *26*(3), 303.

Roh, M. C., Fazli, S., & Lee, S. W. (2016). Selective temporal filtering and its application to hand gesture recognition. *Applied Intelligence*, *45*(2), 255–264.

Saengsri, S., Niennattrakul, V., & Ratanamahatana, C. A. (2012). TFRS: Thai finger-spelling sign language recognition system. In *2nd International Conference on Digital Information and Communication Technology and Its Applications, DICTAP* (pp. 457–462). Bangkok, Thailand.

Sahoo, A. K., Mishra, G. S., & Ravulakollu, K. K. (2014). Sign language recognition: state of the art. *ARPN Journal of Engineering and Applied Sciences*, *9*(2), 116–134.

Sawant, S. N. (2014). Sign Language Recognition System to aid Deaf-dumb People Using PCA. *International Journal of Computer Science and Engineering Technology*, *5*(5), 570–574.

Simpson, P. K. (1992). Fuzzy min-max neural networks. I. Classification. *IEEE Transactions on Neural Networks*, *3*(5), 776–786.

Starner, T., & Pentland, A. (1997). Real-time american sign language recognition from video using hidden markov models. In *Motion-Based Recognition* (pp. 227–243). Dordrecht: Springer.

Wang, H., Leu, M. C., & Oz, C. (2006). American Sign Language Recognition Using Multi-dimensional Hidden Markov Models. *Journal of Information Science and Engineering*, *22*(5), 1109–1123.

Wang, R. Y., & Popović, J. (2009). Real-time hand-tracking with a color glove. In *ACM transactions on graphics (TOG)* (Vol. 28, p. 63). New Orleans, Louisiana.

Yan, S., Xia, Y., Smith, J. S., Lu, W., & Zhang, B. (2017). Multiscale Convolutional Neural Networks for Hand Detection. *Applied Computational Intelligence and Soft Computing*, *2017*, 1-13.

Yang, W., Tao, J., Xi, C., & Ye, Z. (2015). Sign Language Recognition System Based on Weighted Hidden Markov Model. In *8th International Symposium on Computational Intelligence and Design (ISCID)* (pp. 449–452). Hangzhou, China.

Zabulis, X., Baltzakis, H., & Argyros, A. (2009). Vision-based hand gesture recognition for human-computer interaction. In *The universal access handbook* (pp. 1–30). Boca Raton: CRC Press.

Zaki, M. M., & Shaheen, S. I. (2011). Sign language recognition using a combination of new vision based features. *Pattern Recognition Letters*, *32*(4), 572–577.

Zhang, L. G., Chen, Y., Fang, G., Chen, X., & Gao, W. (2004). A Vision-based Sign Language Recognition System Using Tied-mixture Density HMM. In *Proceedings of the 6th international conference on Multimodal interfaces* (pp. 198–204). State College, Pennsylvania, USA.