Real-Time Sign Language Learning System

Lester Wong Sze Ee*, Chandra Reka Ramachandiran and Rajasvaran Logeswaran

Asia Pacific University of Technology and Innovation, Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Malaysia

*Corresponding author e-mail: lesterwong97@gmail.com

Abstract. Learning sign language can be fun and easy when a suitable learning method is used. Different people would have different interest level in learning sign language, and this also can be affected by their communication skills. With the traditional teaching method, there would be a lot of constraints that could stop the learner from learning effectively. This work proposes a learning platform to address these concerns and facilitating sign language learners in gaining the required knowledge. The developed web application includes online learning materials such as videos, and incorporates interactive elements such as a practice session to increase the efficiency of the learning process. The findings indicate that the Sign Language Learning System could be effective for those who are interested in learning sign language.

Index Terms. Sign language, Hearing-impaired communication, Online skills education.

1. Introduction

People express their feelings and much of the body language commonly by the use of hand gestures, which could help them express themselves more convincingly [1]. Sign language is based on hand gestures and used by those who are hearing- and/or verbally-challenged for daily interactions and communication. The World Health Organization (WHO) had identified that more than 5% of the population of the world has hearing loss, including children. An estimated one out of ten people in the world will have hearing disability by 2050 [2].

There are two types of sign language, which are signs and finger spellings. Signs are dynamic gestures which use non-stop hand motions and different hand gestures, while finger spellings are static gestures that apply different hand configurations to represent the letters of the alphabet [3]. Finger spelling is the process of spelling out words by using hand shapes, which is known as a "manual alphabet." Each sign gesture combined with movement of the hands could have a special meaning. These can also include facial expressions and body movement to bring out different meaning. However, different sign languages may be used across different countries. For example, the American Sign Language (ASL) is used in America, British Sign Language (BSL) is used in England and others in different countries [4]. Alternative methods such as written communication could be cumbersome, impersonal and even impractical when an emergency occurs [5].

The importance of learning sign language often be ignored because there are no disable people around us. Therefore, a communication barrier that separates the normal and disabled communities is formed.
Fortunately, rapid development of information technology (IT) and human–computer interaction (HCI) technology enables us to interact with computers by getting feedback, which has given rise to many sign recognition systems [6], providing opportunities for users to learn sign language faster and easier. In [7], Parvez state that developing communication or linguistic skills can be challenged when the individual are not comfortable with conventional education method. Therefore, it is important to integrate technology into education to make learning experience better.

Sign recognition can be categorized into contact-based and vision-based approaches. The most effective solution is contact-based approach, which involves physical interaction with sensing devices such as data gloves. This is because data gloves could provide a more complete set of information of the hand as it is directly worn by the user. However, besides the fact that they are still very expensive for daily usage, they require complex calibration procedures that involve advance knowledge [8].

Vision-based approach is another way for sign gesture recognition that uses a camera to collect data in the form of images or video frames. Vision-based approach is often preferred over contact-based approaches as this method is natural, easy, and the user could directly interact with the system [9]. However, this approach does face many challenges, such as complex background, variation in lighting condition, color of skin and so on [4] in learning sign language.

Due to the drawbacks in the traditional learning method and lacking of interactive sign language learning applications [10], this study aims to develop a web-based sign language learning system that provides Malaysian Sign Language libraries that accurately encode and identify sign language gesture performed by users using camera.

2. Materials and Methods

This project is based on vision-based approach by utilizing web camera to recognize sign language performed by user and identify the correctness of their gesture by using selected image classification algorithm with custom sign gesture dataset.

2.1. Proposed Methodology

Rational Unified Process (RUP) is one of the iterative software development methodologies that assigns tasks and responsibilities among the development organization by using a disciplined approach. The purpose of this methodology is to build high-quality software that satisfy its end-users requirements, within predictable time and cost [11]. The RUP consists of four consecutive phases, which are Inception, Elaboration, Construction and Transition, as depicted in the Fig. 1. A description of these phases along with their deliverables, are given below.

![Fig.1 Development Phase of Rational Unified Process [11]](image)

2.1.1 Inception
In Inception phase, the aim and objective, scope and business case for this project is established, and all major project risks are identified and analyzed. This phase is to ensure that the project is both valuable and feasible to be accomplish.

2.1.2 Elaboration
Once the project is identified to be feasible and requirements are accepted, the software development plan is established in the elaboration phase to provide stable foundation for design and implementation. Besides that, some programming iteration will be done to deliver the high-risk functionalities and confirm whether these major technical risks can be overcome [12].

2.1.3 Construction
If the project is still feasible to be developed after undergoes elaboration phase, the project will move into construction phase where coding and implementation are taking place. The features of the system will be done incrementally until final system is developed. The end-product of this phase is a beta version which is ready to be deployed and tested in user environment using user acceptance testing [13]. Before that, the system will be going through unit testing and integration testing to make sure the system is working well for every functionality.

2.1.4 Transition
In the transition phase, the system is evaluated and refined based on feedback from the users to increase user satisfaction on the product. New releases will be developed in order to correct some arise problems or to finish the functionality that were postponed from Construction phase [14].

2.2 Sign Language Recognition System
Sign Language Recognition system has been an evolving area of research in computer vision for over a decade [15]. Different techniques with different devices have been proposed to increase and improve accuracy of hand gesture recognition. For example, with the vision-based approach, there are techniques such as image processing, image extraction, image segmentation and object detection. However, choosing a suitable technique is important to achieve the objective of this project, which is to develop an interactive cost-effective sign language recognition system.

2.2.1 Data Acquisition
Acquisition of images is the first step in image processing. Camera is the main device used in the proposed Sign Language Recognition system. Besides camera, other sensor devices such as Microsoft Kinect is also popular in capturing images. Kinect can provide depth data that facilitates segmentation of the background [16]. In [17], deep learning with Microsoft Kinect technology was used to do dynamic hand gesture recognition. Despite all the benefits provided, it comes at a cost that is pricey. This project utilizes a simple camera as the input device to capture the image because using external sensor devices can be cumbersome. Many early SLR systems were based on contact-based approach that used data gloves as input to the SLR. The benefits of the sensor input were that feature extraction was bypassed and the measurements could be used directly as features. Although this technique gives accurate positions, it does not allow full natural movement and the mobility of the signer is restricted. Besides that, the prohibitive costs of this approaches had caused the vision approach to be preferred [18].

2.2.2 Transfer Learning
Transfer learning is an important technique in machine learning to solve the limited training data problem. It transfers the knowledge from the source domain to the target domain by relaxing the assumption that the training data and the test data must be identically distributed [19]. Transfer learning also takes less time to build accurate models since it takes features learned from solving a previous problem to solve a new problem [20]. A pre-trained model is trained on a large amount of similar dataset to solve a similar problem. There are several pre-trained models available such as VGG, Inception and MobileNet [20]. It is preferred that the
A pre-trained model be used with the new network directly rather than retraining the model again due to the high computational cost that may take up to weeks. MobileNet was chosen for the proposed system as it is small in terms of file size, fast, accurate and easy to tune. MobileNets are a family of convolutional neural network (CNN) developed by Google. It primarily focuses on optimizing latency with small networks to be resource-friendly and execute quickly.

2.2.3 **K-Nearest Neighbors**
The k-Nearest Neighbors (KNN) algorithm is a supervised algorithm that classifies objects based on feature space (feature similarity) [21]. It is one of the most popular classification techniques, is easy to implement and can perform quite complex classification tasks.

KNN is a lazy learning algorithm as it accepts and keeps all the data until it receives a request to classify an unlabeled data. This characteristic makes it adaptive to new and small datasets. Moreover, KNN is also non-parametric as it does not assume the underlying data distribution, meaning that the model structure is determined from the input data [22]. KNN is often used for classification where an object is classified to the class by a majority vote of its \( k \) nearest neighbors with similar feature to include in most of the voting process. Therefore, parameter tuning for the \( k \) value is important as it increases prediction accuracy [23]. In [24], a finger-spelling recognition method by using a KNN classifier was proposed. By using histogram as the image pre-processing technique combined with Principal Component Analysis (PCA) to extract features from the image possessed and classified with KNN algorithm, the system successfully achieved an accuracy of 99.8% in detecting finger spelling gestures.

2.3 **Similar Systems**
There are some similar systems related to learning recognition of sign language and sign language. These systems are reviewed for their sign language content and concept.

2.3.1 **Motion Savvy**
Motion Savvy is a tablet app that understands sign language by using Leap motion technology to translate ASL into English and vice versa. Fig. 2 shows the hardware for the system. The software also has voice recognition, which converts the voice into text that the hearing-impaired receiver can understand [25]. The software components of the system is shown in Fig. 3.
2.3.2 Eddy: Digital Learning of Sign Language
EDDY is a mobile application that aims to help make it easy for the deaf community, teachers, family and society to learn the Malaysian Sign Language, or Bahasa Isyarat Malaysia (BIM). EDDY localizes and gamifies signing to provide a more fun and easy way of learning, especially children [26].

2.3.3 System Review
The weakness of the existing sign language learning systems and sign language recognition systems is that they interactivity, mostly limited to videos and pictures. It is important for learners to have the interest to practice the lessons in order to develop their linguistic skill. Therefore, a sign language learning platform that integrates with sign language recognition system for interactive practice is proposed. This interactive practice session is real-time and provides instant feedback on the correctness of sign gesture performed by the learner.

2.4 Dataset
A custom dataset was used to train the Sign Language Recognition System. For the purposes of testing, the system was limited to static finger spelt words. Each character is classed with their title and there are numbers of images in each class. The dataset for the system is as shown in Fig. 4.
3. Results and Discussion

The web-based interface via a browser and webcam captures and displays the gesture performed by the user (see Fig. 5). The image taken is then fed into MobileNet, to get its intermediate activation layer, which is known as ‘conv-preds’. By going through this process, the number of samples needed in each class is greatly reduced as a pretrained model uses high level abstractions that the neural network has learned to recognize the classes. After that, this intermediate activation layer is passed to the KNN classifier that allows to train new classes.

| Character | 1st Attempt | 2nd Attempt | 3rd Attempt |
|-----------|-------------|-------------|-------------|
| A         | 95%         | 90%         | 85%         |
| B         | 100%        | 75%         | 70%         |
| C         | 100%        | 90%         | 75%         |

Table 1. Accuracy of Sign Language Recognition System

For its result, the system is tested with 3 attempts for 20 times of fingerspelling recognition. The first attempt is used to test whether the system detects wrong fingerspelling gestures. The second attempt tests the accuracy of different angles of the gestures. The third attempt is tested with different environment. Table 1 shows the accuracy of recognizing different characters.
Based on the results, the accuracy of the system was satisfactory. However, the lighting and complexity of the background did affect the accuracy. Under good lighting environment and light-colored background, the accuracy of the system could achieve nearly 100% accuracy.

4. Conclusions
This web application was successfully developed and built via Node.JS with Express.js as its framework. This project managed to gain satisfactory results but could be further improve in terms of invariability to influences of lighting and background as mentioned in [4] and [27]. Further functionalities could also be incorporated. The future direction could include configuring the training of the recognition system to be more flexible and designing a more interactive practice session.

References
[1] M. J. Cheok, Z. Omar and M. H. Jaward, "A review of hand gesture and sign language recognition techniques. International Journal of Machine Learning and Cybernetics," 2017.
[2] who.int, "Deafness and hearing loss," 2019. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss. [Accessed 1 February 2020].
[3] R. G. Rajan and M. J. Leo, "A Comprehensive Analysis on Sign Language," International Journal of Recent Technology and Engineering (IJRTE), vol. 7, no. 6, pp. 749-755, 2019.
[4] P. Pandey and V. Jain, "Hand Gesture Recognition for Sign Language Recognition: A Review," International Journal of Science, Engineering and Technology Research (IJSETR), vol. 4, no. 3, pp. 464-470, 2015.
[5] B. Garcia and S. A. Viesca, "Real-time American Sign Language Recognition with Convolutional Neural Network," pp. 225-232, 2016.
[6] T. W. Chong and B. G. Lee, "American Sign Language Recognition Using Leap Motion Controller with Machine Learning Approach," Sensors, vol. 18, no. 10, p. 3554, 2018.
[7] K. Parvez, K. Muzafar, I. Javed, T. Muhammad, A. Ahmed, A. Mohammed, A. A. Amer and J. Nadeem, "Measuring Effectiveness of Mobile Application in Learning Basic Mathematical Concepts Using Sign Language," Sustainability, vol. 11, no. 1, pp. 1-20, 2019.
[8] T. Cardoso, J. Delgado and J. Barata, "Hand Gesture Recognition towards Enhancing Accessibility," 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Infoexclusion (DSAI 2015), pp. 419-429, 2015.
[9] S. Padmavathy, M. Nellaiappan and R. L. J. Femila, "Gesture Recognition Techniques," IJARIIE, vol. 3, no. 4, pp. 3128-3838, 2017.
[10] S. A. Mokhtar and S. M. S. Anuar, "Learning application for Malaysian sign language: content design, user interface and usability," Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication - IMCOM'15, 2015.
[11] T. K. Tia, I. Nuryasin and Maskur, "Simulation Model For Rational Unified Process (RUP) Software Development Life Cycle," Jurnal SISTEMASI, vol. 8, no. 1, pp. 176-184, 2019.
[12] R. Hughes, "Chapter 3 - Introduction to Alternative Iterative Methods," in Agile Data Warehousing for the Enterprise: A Guide for Solutions Architects and Project Leaders, 2016, pp. 31-54.
[13] Info-Novitas, "Rational Unified Process (RUP)," 2019. [Online]. Available: https://en.info-novitas.hr/about-us/work-methodologies/rup/. [Accessed 10 January 2020].
[14] M. Mrsic, "Rational Unified Process (RUP)," 2017. [Online]. Available: https://activecollab.com/blog/project-management/rational-unified-process-rup. [Accessed 10 January 2020].
[15] G. A. Rao and P. V. V. Kishore, "Selfie Sign Language Recognition With Multiple Features On Adaboost Multilabel Multiclass Classifier," Journal of Engineering Science and Technology, vol. 13, no. 8, pp. 2352 - 2368, 2018.
[16] Suharjito, R. Anderson, F. Wiryana, M. C. Ariesta and G. P. Kusuma, "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output," Based on Input-Process-Output. Procedia Computer Science, pp. 441-448, 2017.

[17] Q. Smedt, "Dynamic hand gesture recognition - From traditional handcrafted to recent deep learning approaches," 2017.

[18] H. Cooper, B. Holt and R. Bowden, "Sign Language Recognition," 2011.

[19] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang and C. Liu, "A Survey on Deep Transfer Learning," Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing, China, 2018.

[20] P. Marcelino, "Transfer learning from pre-trained models," 2018. [Online]. Available: https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751. [Accessed 2 January 2020].

[21] P. Ravi, "A Review on Image based Indian Sign Language Recognition," International Journal of Innovative Research in Computer and Communication Engineering, vol. 6, no. 12, pp. 9101-9107, 2018.

[22] S. Swati, P. S. Varun and S. Kandukuru, "Comparative Analysis On Sign Language," International Journal Of Scientific & Technology, vol. 8, no. 8, pp. 981-990, 2019.

[23] D. Subramanian, "A Simple Introduction to K-Nearest Neighbors Algorithm," 2019. [Online]. Available: https://towardsdatascience.com/a-simple-introduction-to-k-nearest-neighbors-algorithm-b3519ed98e. [Accessed 5 January 2020].

[24] D. Aryanie and H. Yaya, "American sign language-based finger-spelling recognition using k-Nearest Neighbors classifier," 3rd International Conference on Information and Communication Technology (ICOICT), pp. 533-536, 2015.

[25] T. Alexia, "MotionSavvy Is A Tablet App That Understands Sign Language," 7 June 2014. [Online]. Available: https://techcrunch.com/2014/06/06/motionsavvy-is-a-tablet-app-that-understands-sign-language/. [Accessed 18 October 2019].

[26] Digi, "Digi: Learn Basic Malaysian Sign Language with EDDY," Eddy, 29 December 2017. [Online]. Available: https://www.malaysianwireless.com/2017/12/digi-malaysian-sign-language-eddy/. [Accessed 10 October 2019].

[27] R. Wario and C. Nyaga, "A Survey of the Constraints Encountered in Dynamic Vision-Based Sign Language Hand Gesture Recognition," Universal Access in Human-Computer Interaction, pp. 373-382, 2019.