NCSE 2019, 27th - 28th June 2019, Yangon

Sign Language Recognition for Myanmar Number Using Three Different SVM Classifiers

[1*] Ni Htwe Aung, [2] Su Su Maung, [3] Ye Kyaw Thu
[1*, 2] Department of Computer Engineering and Information Technology, Yangon Technological University, Yangon, Myanmar
[3] Language and Speech Science Lab, Waseda University, Japan,
[3] Language and Semantic Technology Research Team, National Electronics and Computer Technology Center (NECTEC), PathumThani, Thailand
[1*]nhadec@gmail.com

***Abstract*— People who are affected by hearing problems use a communication method called Sign Language. Sign Language differs from region to region, country to country and continent to continent. Machine learning can play a significant role in impacting lives of the hearing-impaired people and can help them to communicate with their environments more easily. This paper presents a very simple and efficient approach for Myanmar Sign Language (MSL) recognition system which is capable of recognizing the static hand gesture images that represent the Myanmar numbers from zero to ten. The main objective of this paper is to investigate the performance of three different Support Vector Machine (SVM) classifiers for Myanmar number sign recognition. The proposed system contains three stages, namely, pre-processing, feature extraction and classification. In the feature extraction stage, different features are extracted using Scale Invariant Feature Transform (SIFT) algorithm. In the classification stage, three different SVM classifiers (SVCs); SVC with linear kernel, SVC with polynomial kernel and LinearSVC are tested and evaluated. Among these three classifiers, SVC with polynomial kernel yielded the highest accuracy score with 87%. Although there are some limitations in the datasets, each classifier provides the encouraging results.**

Keywords: Myanmar Sign Language, SIFT, SVM, Myanmar number sign

I. INTRODUCTION

In recent years, many researchers have been paying attention to the research area of Sign Language (SL) recognition. It is important for many research fields as computer vision, natural language processing, Human Computer Interaction (HCI), image processing and linguistics. SL recognition system still remains as a challenging task because sign language is a visual language which contains the motion of the body, head, eyes, hands and facial expressions. SLs can differ from region to region based on the culture and environments of these particular regions. Therefore, it cannot be clearly said that how many SLs are used in the world.

In Myanmar, there are 673,126 hearing-impaired persons according to the 2014 Myanmar national census [5]. Myanmar Deaf people use Myanmar Sign Language (MSL) to communicate with other deaf people or hearing people. However, MSL used in southern Myanmar and MSL used in northern Myanmar is also different. Moreover, very little research work can be found in the area of MSL recognition system. To the best of our knowledge, the proposed system would be the first MSL recognition system for Myanmar numbers '၀' (0) to '၁၀' (10). In this paper, the performance of three different SVM classifiers (SVCs); SVC with 'linear' kernel, SVC with 'polynomial' kernel and LinearSVC are investigated using MSL images of Myanmar numbers. Most of the number signs from other countries are one-handed signs, but some of the number signs used in MSL contain two-handed signs. Furthermore, most of the existing approaches developed SL number recognition for only '0' to '9' and number '10' was excluded because it is signed by using two hands. We recorded the MSL videos of Myanmar numbers '၀' (0) to '၁၀' (10) demonstrated by the deaf signers from Mary Chapman School for the deaf, Yangon. These videos were converted into the corresponding image frames and feature vectors of these images are extracted using Scale Invariant Feature Transform (SIFT) algorithm. The extracted features of MSL numbers images are trained and classified by using three different types of SVM classifiers.

II. MYANMAR SIGN LANGUAGE

MSL is the native language and an essential communication tool of the Myanmar deaf people. The grammatical structure of MSL is different from Myanmar Language. As shown in Figure 1, MSL is also constructed with manual and non-manual components like other SLs such as American Sign Language, British Sign Language, etc. Manual components contain hand shape, hand position and hand movement. Non-manual components are used to show feeling and meaning with facial expressions, movement of head, tension, eyebrows, eyelid and body [12]. In Myanmar, there are only four deaf schools that use and teach MSL: Mary Chapman School for the Deaf, Yangon (est. 1904), School for the Deaf (Tamwe, Yangon), School for the Deaf, Mandalay (est. 1964) and Immanuel School for the Deaf, Kalay (est. 2005) [2]. MSLs used in each region are also different. In Myanmar, two MSLs are used in the lower part and upper part differently.

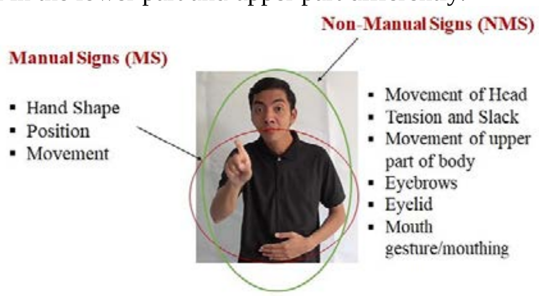


Figure 1. Structure of Myanmar Sign Language (MSL)

III. RELATED WORK

Sign Language recognition system was developed almost 20 years ago and the first publications had emerged in the beginning of the 90s. Additionally, most proposed approaches needed the use of expensive hardware devices such as gloves, 3D camera or low noise and high resolution images. Wah Wah et al. [11] developed first Myanmar Fingerspelling Recognition System for 30 static and opened fingerspelling characters of Myanmar alphabets ‘က’ (ka) to ‘အ’ (a) excluding the closed fingerspelling character for ‘ဏ’ (nna) and the dynamic fingerspelling characters for ‘ဇ’ (ja), ‘ဈ’ (jha). Their approach used five stages in preprocessing, canny edge detection for feature extraction and Artificial Neural Network (ANN) for training and testing and obtained the accuracy of 96%. In our previous work, the Myanmar Fingerspelling Recognition System for 31 static fingerspelling characters of Myanmar alphabets including both opened and closed fingers images were developed [9]. Our system provided the higher accuracy of 97% using Random Forest Classifier.

IV. IMPLEMENTATION

A. Dataset and Pre-processing

For the system implementation, MSL videos for Myanmar numbers have recorded from ‘Mary Chapman School for the Deaf (Yangon)’, using Canon 200D camera with the resolution of 1,920x1,080. All videos were taken in indoor environments using light-yellow background under normal lighting conditions. The upper part of the body of the deaf signers was captured during the video recording periods. And then, the recorded MSL number videos were converted into multiple image frames and from these image frames, the areas of the hand region were segmented manually. The segmented hand-only images are resized into 128x128 images and then converted them into the grayscale images. The dataset of this proposed system consists of 951 one-handed and two-handed images for each of the 11 different hand signs for Myanmar numbers from ‘၀’ (0) to ‘၁၀’ (10). In our proposed system, very few pre-processing stages are required. Figure 2 shows MSL numbers images used in this proposed system and the system flow diagram is displayed in Figure 3.

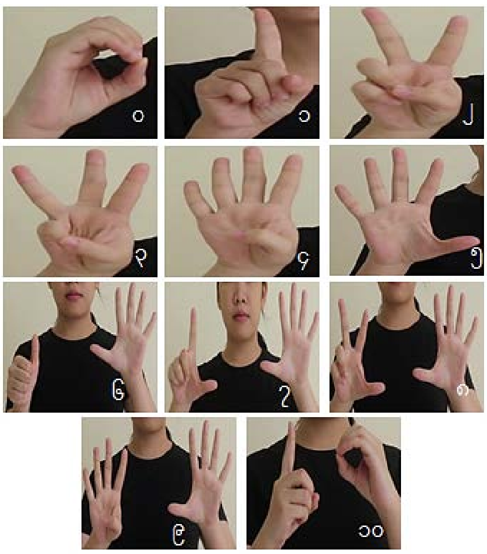


Figure 2. Images of Myanmar signs for Myanmar numbers

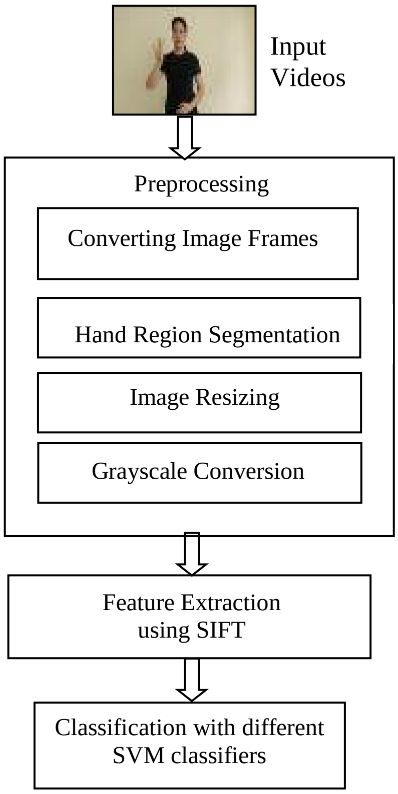


Figure 3. System flow diagram of the proposed system

B. Feature Extraction

Feature extraction is the process of transforming the input raw data into the reduced form to facilitate decision making such as pattern recognition, face detection, classification or recognition [6]. In our proposed system, we extracted 128 feature vectors from each preprocessed image using SIFT algorithm. This algorithm was developed by David G. Lowe, the University of British Columbia in 1999, to extract distinctive invariant features from the image data and compute its descriptors [7] [8]. SIFT transforms an image into a large collection of local feature vectors called SIFT keys, each of which is invariant to image translation, rotation, scaling and illumination changes. An example image of MSL number that uses SIFT is illustrated in Figure 4.

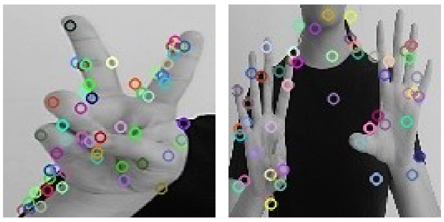
The SIFT algorithm can be divided into four main steps.

1. **Scale-Space Extrema Detection:** This is the first stage where the Gaussian filter is applied to detect the interest points, that are the keypoints in the SIFT framework.

2. **Key point Localization:** In this stage, the keypoints which have high contrast are chosen and the keypoints which have low contrast or are poorly localized on an edge are eliminated to calculate the accurate location and scale.

3. **Orientation Assignment:** This is the third stage that assigns one or more orientations to each keypoint in order to achieve invariance to rotation based on local image gradient directions. A histogram of orientation is established from the gradient orientations of sample points on the image data.

4. **Key point Descriptor:** The computation of descriptor vector for each keypoint is done in this stage. A set of orientation histograms is built around every keypoint and then concatenated in a 128-values vector.



(a) (b)

Figure 4. Sample SIFT image for Myanmar numbers (a) '၂' (two) and (b) '၉' (nine)

C. Classification

Classification is the process of predicting the class of given data. It is also a supervised machine learning approach in which the classifier model is built from the input data and then this model is used to classify the new data. In Machine learning, there are many types of classification algorithms such as Random Forest, k-Nearest Neighbor (KNN), Support Vector Machines and so on. In our proposed system, Support Vector Machines (SVM) is used for the classification process [4]. SVM is a supervised learning algorithm and was introduced initially in 1960s [3]. It is still becoming a popular algorithm because of its ability to achieve brilliant classification results. SVM constructs a hyperplane to separate the data into different classes. The main idea of SVM is to generate the maximum hyperplane margin to divide the dataset into corresponding classes accurately [3].

In practice, SVM algorithm is implemented using a kernel, which transforms an input data space into the required form [10]. The kernel converts the non-separable problem into the separable problems by adding more dimensions to it. In SVM Classifier (SVC), there are many types of SVM classifiers with different kernel functions such as linear, nonlinear, polynomial, Gaussian kernel, Radial Basis Function (RBF), sigmoid etc. [1]. In this proposed system, the performances for three different kernel SVCs; SVC with linear kernel, SVC with polynomial kernel and LinearSVC were investigated.

SVC with linear kernel minimizes the regular hinge loss and uses the One-vs-One multiclass reduction. Polynomial kernel is a generalized form of the linear kernel. The flexibility of the resulting classifier is controlled by the degree of the polynomial kernel. For our proposed system, we used degree 8 for the polynomial kernel. Linear SVC is similar to SVC with linear kernel but slightly different in decision boundaries. It minimizes the squared hinge loss and uses the One-vs-All multiclass reduction. The performance results for these three classifiers are compared and discussed in Section V.

V. Experimental Results

The Myanmar number SL images for 11 different classes are trained and tested using SVM classifiers with three different kernels. In our experiment, 713 images were used for the training phase and 238 images were used for testing the model. Accuracy score for linear kernel is 79%, 87% for polynomial kernel and 77% for LinearSVC. The graph of the confusion matrix for each classifier is illustrated in Figure 5, 6, and 7 to easily understand the performance of each classification model. The classification result for each classifier is also shown graphically in Figure 8.

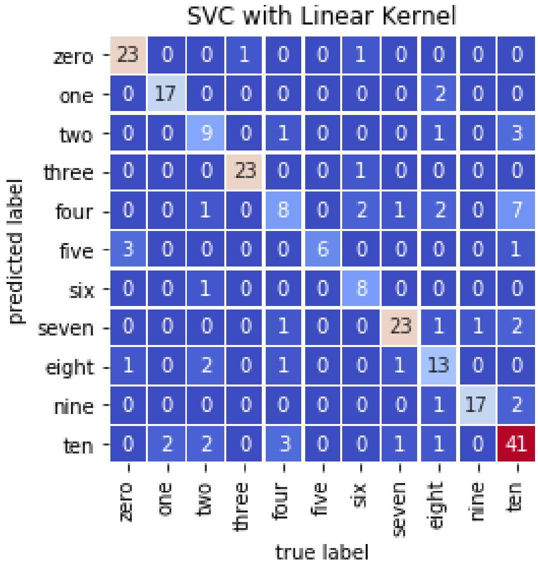


Figure 5. Confusion matrix of SVM with linear kernel

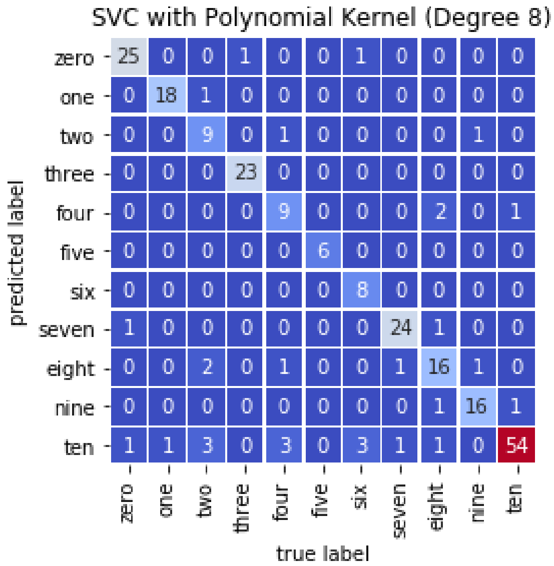


Figure 6. Confusion matrix of SVM with polynomial kernel (degree 8)

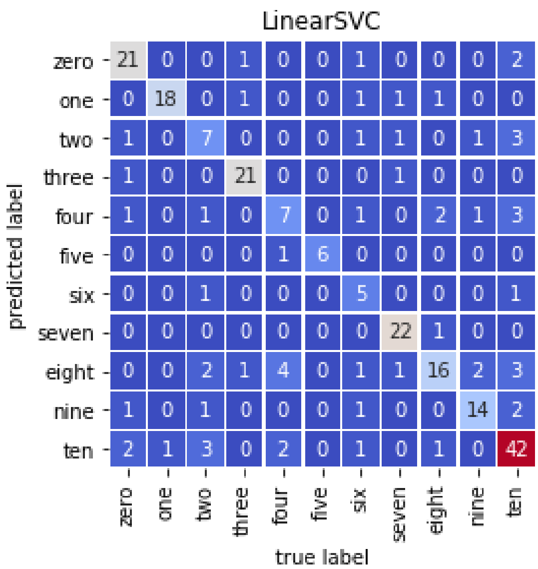


Figure 7. Confusion matrix of LinearSVC

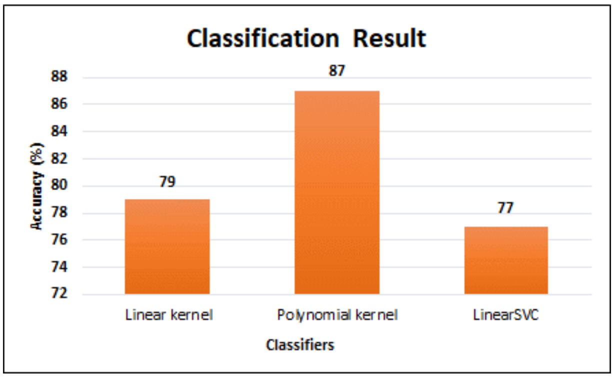


Figure 8. Classification result of each classifier

TABLE I. CROSS VALIDATION SCORE

Classifier	Linear Kernel	Polynomial Kernel (degree 8)	LinearSVC
10-fold Cross Validation Score	84%	92%	72%
	80%	84%	82%
	80%	88%	74%
	84%	90%	80%
	86%	88%	81%
	85%	94%	80%
	81%	82%	64%
	81%	88%	85%
	83%	88%	84%
	86%	89%	84%
Average	83%	**88%**	79%

In this experiment, we made 10-fold cross validation for each classifier to estimate the performance of each model on new data. The results of 10-fold cross validation for each classifier are shown in Table 1. The highest average result among three classifiers is highlighted. According to this table, it can be seen that the best accuracy score of 94% was achieved in one round of 10-fold cross validation for the classifier with polynomial kernel.

VI. CONCLUSION

Although there are some limitations in the datasets in our experiment, we obtained the encouraging result with very few image preprocessing stages. Moreover, our system is capable of classifying 11 Myanmar Number characters for MSL without the need for any special expensive hardware devices such as gloves, 3D cameras or sensors. In future, we intended to develop a Myanmar Fingerspelling recognition system for all Myanmar consonants, vowels and symbols including static and dynamic signs by applying deep neural network.

ACKNOWLEDGMENT

We would like to thank principals, teachers, SL translators and students from Mary Chapman School for the Deaf (Yangon), School for the Deaf, Tamwe (Yangon) and all participants for their kind contributions to our research. We would like to thank JICA EEHE Project (Project for Enhancement of Engineering Higher Education in Myanmar) for their supporting of research fund to our research.

REFERENCES

[1] Apostolidis-Afentoulis, Vasileios, vapostolidis@gmail.com, “SVM Classification with Linear and RBF kernels”, 10.13140/RG.2.1.3351.4083, (2015).

[2] Burmese Sign Language (or) Myanmar Sign Language, https://en.wikipedia.org/wiki/Burmese_sign_language

[3] Chaitali Dhaware, Mrs. K. H. Wanjale, “Survey On Image Classification Methods In Image Processing”, International Journal of Computer Science Trends and Technology (IJCST) – Volume 4 Issue 3, May - Jun 2016

[4] Cristianini, Nello; and Shawe-Taylor, John; “An Introduction to Support Vector Machines and other kernel-based learning methods”, Cambridge University Press, 2000. ISBN 0-521-78019-5

[5] Disability population in Myanmar, 2014

[6] Gaurav Kumar, Panipat, Haryana, India, er.gkgupta@gmail.com, Pradeep Kumar Bhatia, Hisar, Haryana, India, pkbhatia.gju@gmail.com, “A Detailed Review of Feature Extraction in Image Processing Systems”, Fourth International Conference on Advanced Computing and Communication Technologies, 2014

[7] Lowe, David G. (1999). “Object recognition from local scale-invariant features”, Proceedings of 7th International Conference on Computer Vision (ICCV'99) (Corfu, Greece): 1150-1157.

[8] Lowe, David G. “Distinctive Image Features from Scale-Invariant Keypoints”, Computer Science Department, University of British Columbia, Vancouver, B.C., Canada, lowe@cs.ubc.ca, International Journal of Computer Vision 60(2): 91-110, January 5, 2004

[9] Ni Htwe Aung, Ye Kyaw Thu, Su Su Maung, “Feature Based Myanmar Fingerspelling Image Classification Using SIFT, SURF and BRIEF”, In Proceedings of the 17th International Conference on Computer Applications (ICCA, 2019), pp. 245-253, February 2019.

[10] Theodoridis, Sergios; and Koutroumbas, Konstantinos; Pattern Recognition, 4th Edition, Academic Press, 2009, ISBN 978-1-59749-272-0

[11] Wah Wah, “Myanmar Sign Language Recognition System Using Artificial Neural Network”, December, 2014

[12] Yu Yu Swe, Department of Social Welfare, "Myanmar Sign Language Basic Conversation Book", Ministry of Social Welfare, Relief and Resettlement, Department of Social Welfare, Japan International Cooperation Agency, 1st Edition, August 2009,