International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 5, August 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Real-time Myanmar Sign Language Recognition System using PCA and SVM

Myint Tun, Thida Lwin

Department of Information Technology, Technological University, Pathein, Myanmar

How to cite this paper: Myint Tun | Thida Lwin "Real-time Myanmar Sign Language Recognition System using PCA and SVM" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.2361-2366, **https://doi.org/10.31142/ijtsrd26797**

IJTSRD26797

ABSTRACT

Communication is the process of exchanging information, views and expressions between two or more persons, in both verbal and non-verbal manner. The sign language is a visual language used by the people with the speech and hearing disabilities for communication in their daily conversation activities. Myanmar Sign Language (MSL) is the language of choice for most deaf people in this country. In this research paper, Real-time Myanmar Sign Language Recognition System (RMSLRS) is proposed. The major objective is to accomplish the translation of 30 static sign gestures into Myanmar alphabets. The input video stream is captured by webcam and is inputed to computer vision. The incoming frames are converted into YCbCr color space and skin like region is detected by YCbCr threshold technique. The hand region is also segmented and converted into grayscale image and morphological operation is applied for feature extraction. In order to translate the signs of ASL into the corresponding alphabets, PCA is used for feature extraction and SVM is used for recognition of MSL signs. Experimental results show that the proposed system gives the successful recognition accuracy of static sign gestures of MSL alphabets with 89%.

KEYWORDS: *YCbCr Color Space, Threshold, Computer Vision, Feature Extraction, PCA, Real-time, SVM, Sign Language*

I. INTRODUCTION

A Sign Language (SL) means using gestures instead of sound or spoken words to convey meaning combining hand-shapes, orientation and movement of the hands, arms or body, facial expressions and lip-patterns. Gestures can also be divided into static gestures and dynamic gestures. A static gesture is determined by a certain configuration of the hand, while a dynamic gesture is a moving gesture determined by a sequence of hand movements and configurations. In MSL, there are 30 static gestures and 3 dynamic gestures.

In the proposed system, input video stream is inputed computer vision blob analysis vision. Then incoming frames of the stream are checked whether the frame involves skin. The skin-like regions detection and rejection other background regions are done by using YCbCr thresholding technique. When hand gesture region is detected, it is necessary to use morphological operations. The morphological close and open operation is applied. Moreover, holes filling operation is also needed to remove non hand regions. Then the background is removed and the hand region is bounded by a rectangle. After bounding hand gesture by a rectangle, this hand region is cropped. This cropped region is converted into grayscale image and it is filtered by median filter. After that, the features of hand gesture are extracted using Principal Component Analysis (PCA). After getting features from hand gesture, the feature vector is inputed to the Support Vector Machine (SVM). The SVM classifies the hand gestures.

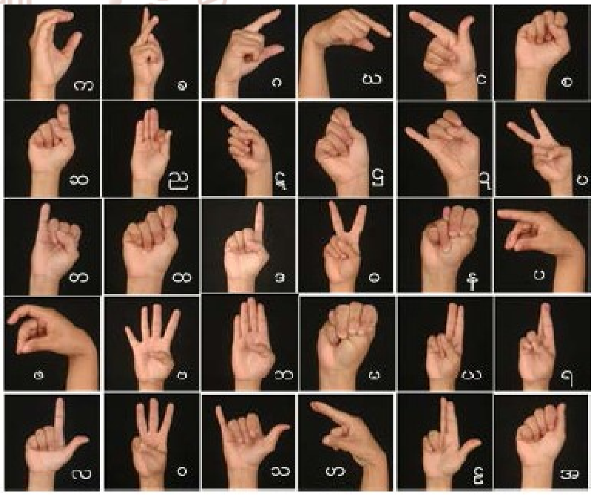


Fig.1.1. Static Myanmar Signs

This research work uses 3000 hand gesture images for 30 static alphabets from different ten signers. The features of the hand gestures are extracted using PCA and SVM is used for training and recognition. Moreover, the proposed system is implemented by using MATLAB programming language for the entire thesis.

II. RELATED WORKS

SL recognition is not a new computer vision problem. Over the past two decades, researchers have used classifiers from a variety of categories that can be grouped roughly into linear classifiers.

L. Pigou et al's application of CNN's to classify 20 Italian gestures from the ChaLearn 2014 Looking at People gesture spotting competition [1]. They use a Microsoft Kinect on full-body images of people performing the gestures and achieve a cross-validation accuracy of 91.7%. As in the case with the aforementioned 3-D glove, the Kinect allows capture of depth features, which aids significantly in classifying ASL signs.

Singha and Das obtained accuracy of 96% on 10 classes for images of gestures of one hand using Karhunen-Loeve Transforms[2]. These translate and rotate the axes to establish a new coordinate system based on the variance of the data. This transformation is applied after using a skin filter, hand cropping and edge detection on the images. They use a linear classifier to distinguish between hand gestures including thumbs up, index finger pointing left and right, and numbers.

Liang, R. H. and Ouhyoung, M. [3] developed real time continuous gesture recognition of sign language using a Data Glove. First, it solved the end-point detection in a stream of gesture input and then statistical analysis is done according to 4 parameters in a gesture: posture, position, orientation, and motion. This system implements a prototype system with a lexicon of 250 vocabularies in Taiwanese Sign Language (TWL). Hidden Markov Models (HMM) can be continuously recognized in real-time and the average recognition rate is 80.4%.

Christopher, L. and Xu, Y. [4] developed a glove-based gesture recognition system based on Hidden Markov Models (HMM). CyberGlove is used to recognize gesture from sign language alphabet.HMM can interactively recognize gestures and perform online learning of new gestures. It can also update its model of a gesture iteratively with each example it recognizes. This system can be used to make cooperation between robots and humans easier in applications such as teleoperation and programming.

Suk et al. proposed a method for recognizing hand gestures in a continuous video stream using a dynamic Bayesian network (DBN) model [5]. They attempt to classify moving hand gestures, such as making a circle around the body or waving. They achieve an accuracy of over 99%, but it is worth noting that all gestures are markedly different from each other and that they are not American Sign Language. However, the motion-tracking feature would be relevant for classifying the dynamic letters of ASL: j and z.

Kadous, M. W. [6] employed instance-based learning and decision tree methods on Australian Sign Language data which are collected by a Power glove. The user dependent system could recognize a set of 95 isolated signs. The accuracy of the proposed method is about 80%. Akyol and Canzler [7] used color coded gloves for finding and tracking the hands reliably. They obtained a user independent recognition rate of 94% on a set of 16 signs of German sign language.

Grobel, K. and Assan, M. [8] used HMM to recognize isolated signs with 91.3% accuracy out of a 262-sign vocabulary. They extracted 2D features from video recordings of signers wearing colored gloves.Cavallo, A. He used glove where 18markers are attached with it, of which 15 are for fingersand three for the reference taken. The image captured is thenclassified based on Singular Value Decomposition (SVD).

Assaleh, et al. [9] proposed a low complexity classification method to recognize Arabic sign language using sensor based gloves. The gloves have five bend sensors and a 3D accelerometer. Their system yields recognition rates of 92.5% and 95.1% for user dependent and user independent cases, respectively.

Grimes, G. J. [10] proposed the idea that a glove equipped with sensors and associated electronic logic can be designed for manual sign recognition in American sign language. He described the overall framework as follows: a deaf or hearing impaired user wearing the data entry glove to input data performs a manual sign and the receiving device converted the signal into a single sign in the finder spelling alphabet.

III. BACKGROUND THEORY

RMSLRS is a vision based recognition system. In this system, object detection, skin detection, feature extraction and classification process are included.

A. Viola-Jones Object Detection Framework

The Viola–Jones object detection framework is the first object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola and Michael Jones. To detect vision based object in real time, this algorithm is used.

B. YCbCr Color Space

The YCbCr color space consists of the Y channel, which represents the luminance component, and the Cb and Cr channels, which describe the chrominance components. The separation of luma from chromatic makes this color model attractive for skin color detection. The separation of brightness information from the chrominance and chromaticity in YCbCr color space can reduces the effect of uneven illumination in an image [11]. This color space is robust against of skin. Therefore, YCbCr are typically used in color conversion.

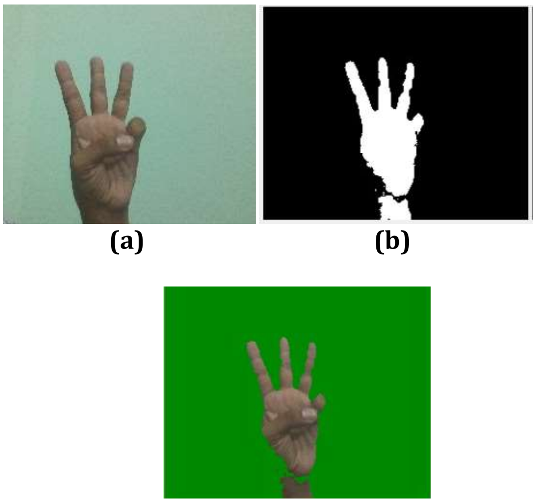


Fig.2. (a) Input Image (b) Skin Detection and (c) Background Removal

The YCbCr color space has good performance of clustering, and the transformation from RGB to YCbCr is linear. More importantly, different from RGB color space, the chrominance component and luminance component in YCbCr color space are separated explicitly, so it is appropriate for skin color segmentation. Y indicates the luminance component and it is the weighted sum of RGB values. Cr and Cb are the red difference and the blue-difference chrominance components. The conversion formula of transformation from RGB to YCbCr is as follows:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.256 & 0.5041 & 0.0979 \\ -0.1482 & -0.291 & 0.4392 \\ 0.4392 & -0.3677 & -0.0714 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

A binary image is got after threshold segmentation. For better performance, it is necessary to remove background and mark the hand region. So, the morphological close and open operation is applied. Moreover, holes filling operation is also needed to remove non hand regions.

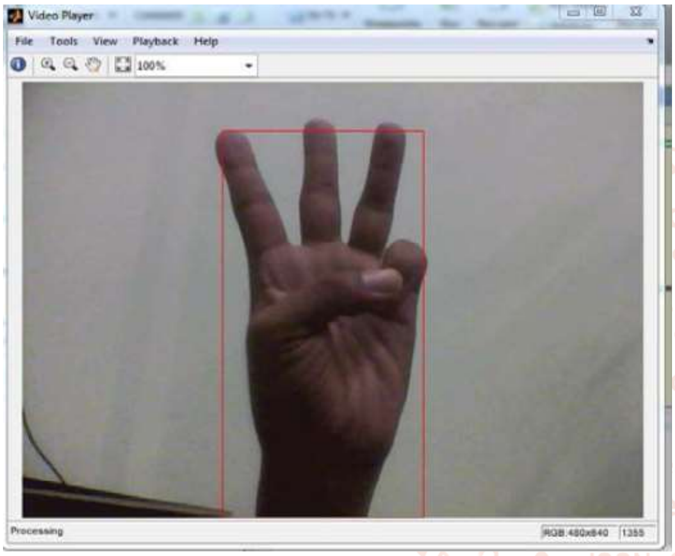


Fig.3. Hand region is bounded by rectangle box

C. Principal Component Analysis (PCA)

PCA is the general name for a technique which uses sophisticated underlying mathematical principles to transforms a number of possibly correlated variables into a smaller number of variables called principal components [12]. PCA is a statistical method under the broad title of factor analysis. The purpose of PCA is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically. This is the case when there is a strong correlation between observed variables. The jobs which PCA can do are prediction, redundancy removal, feature extraction, data compression, etc. Because PCA is a classical technique which can do something in the linear domain, applications having linear models are suitable, such as signal processing, image processing, system and control theory, communications, etc. Face recognition has many applicable areas. Moreover, it can be categorized into face identification, face classification, or sex determination [13]. The mathematical formulations of PCA are described below.

1. Mean and Variance

$$Mean(X') = \frac{1}{n} \sum_{i=1}^{n} Xi$$

2. Standard Deviation

$$SD = \sqrt{\frac{1}{n} (\sum_{i=1}^{n} (Xi - X')^2}$$

3. Covariance

$$Cov(X, Y) = \frac{\sum_{i=1}^{n} (Xi - X')(Yi - Y')}{n - 1}$$

4. Eigen Values and Eigen Vectors

$$[A - \lambda I]X = 0$$

D. Support Vector Machine (SVM)

Classification between the objects is easy task for humans but it has proved to be a complex problem for machines. SVM is a supervised machine learning algorithm which can be used for either classification or regression analysis. The basic concept of SVM is to transform the input vectors to a higher dimensional space by a nonlinear transform, and then an optical hyperplane that separates the data, can be found.

The hyperplane should have the best generalization capability. In many cases, the data cannot be separated by a linear function. The use of a kernel function becomes essential in these cases.Three admissible kernel functions are

- Polynomial kernel of degree h
- Gaussian radial basis function kernel
- Sigmoid kernel

Gaussian radial basis function kernel for training and classification will make prediction accuracy highest. The form of kernel function is as follows:

$$K(x, y) = \exp\left[\frac{-\|x - y\|^2}{2\sigma^2} \right]$$

where:
||x-y||^2 is the squared Euclidean distance between the two feature vectors.
σ is the free parameter.

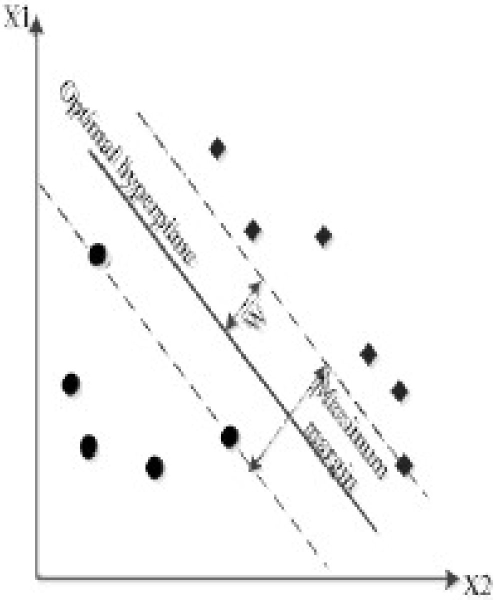


Fig.4. Determination of the Optimum Hyperplane

SVM is designed to solve a binary classification problem. However, for a RMSLR problem, which is a multiple classification problem, classification is accomplished through combinations of binary classification problems. There are two ways to do that: one versus one or one versus all.

Finally, a SVM will be trained for each class, and SVMs of all classes are combined to form a multiclass SVM.

IV. DESIGN OF THE PROPOSED SYSTEM

The flow chart of the proposed system is shown in Figure 1. The input video stream is inputed to computer vision. The video frames are converted to YCbCr color space. Then the system detects skin using YCbCr thresholding technique. Hand region is bounded by rectangular box. In third step, closed morphology operation and bounding box characteristic (height, width, area) are used for all potential regions. The traffic sign is detected and then cropped from the original image. Then it is cropped and preprocessed. After that features are extracted using PCA and classify that feature by SVM.

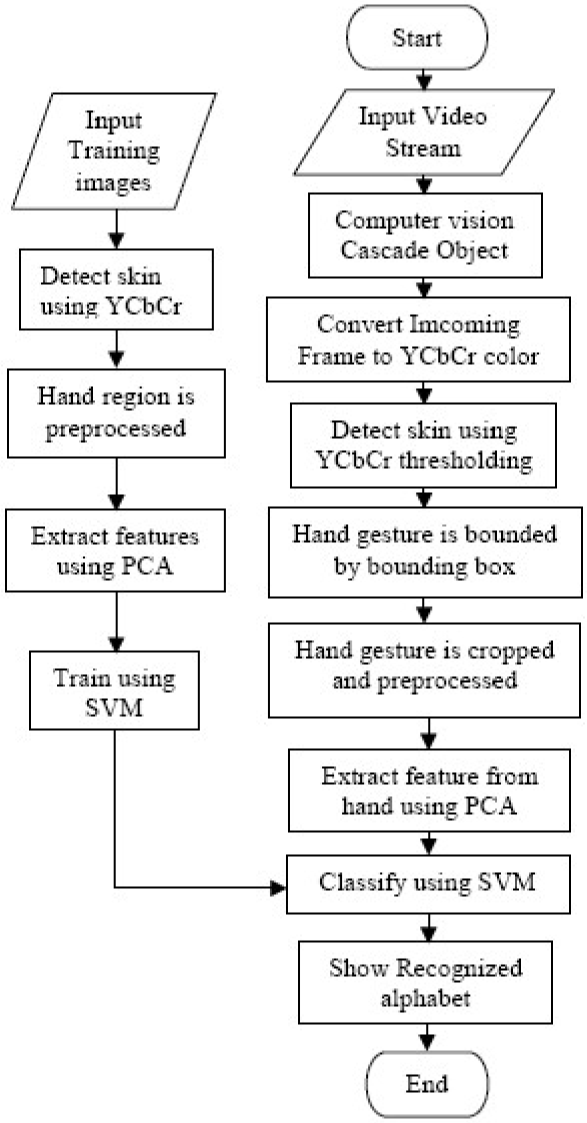
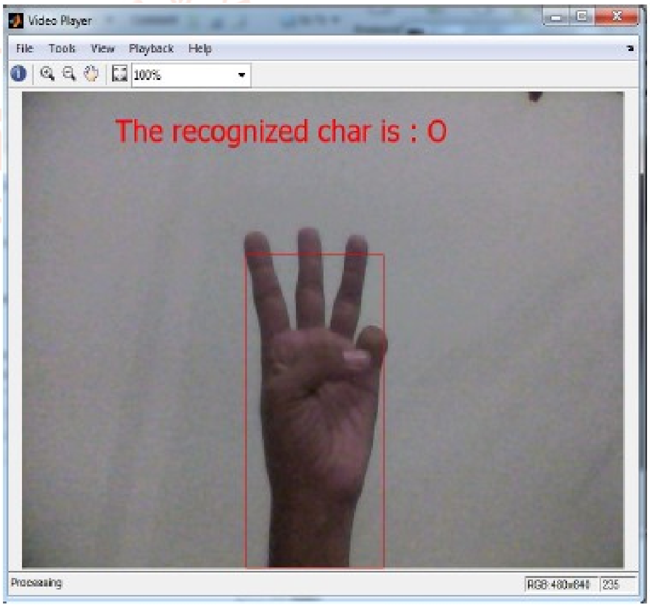

Start
Input Training images
Detect skin using YCbCr
Hand region is preprocessed
Extract features using PCA
Train using SVM
Input Video Stream
Computer vision Cascade Object
Convert Imcoming Frame to YCbCr color
Detect skin using YCbCr thresholding
Hand gesture is bounded by bounding box
Hand gesture is cropped and preprocessed
Extract feature from hand using PCA
Classify using SVM
Show Recognized alphabet
End

Fig.5. The Flow Chart of the Proposed System.

V. EXPERIMENTAL RESULT AND DISCUSSION

The proposed system can handle different types of Myanmar alphabet signs in a common vision-based platform. This system focuses on image processing as a tool for the conversion of MSL gesture into text.

In this proposed system, the data set itself consists of 10 repetitions of each of the 30 alphabets performed by different ten signers, which are used for training. The hand gestures are tested in real time for each alphabet. There are many similar gestures in MSL. Alphabet 'ဂ' is misclassified with 'ဃ', alphabet 'စ' is misclassified with 'ဆ', alphabet 'င' is misclassified with 'လ', alphabet 'ရ' is misclassified with 'ယ', alphabet 'ပ' is misclassified with 'ဖ', alphabet 'ဓ' is misclassified with 'လ' and alphabet 'ဍ' is misclassified with 'တ' because these two hand gestures representation are similarities in shape. And hand gesture representations for alphabet 'ဌ', 'န', 'ထ', 'မ' and 'အ' are also similar in shape. Moreover, Start Input Video Stream Convert Imcoming Frame to YCbCr color Detect skin using YCbCr thresholding Hand gesture is bounded by bounding box Hand gesture is cropped and preprocessed Extract feature from hand using PCA End Input Training images Detect skin using YCbCr Hand region is preprocessed Extract features using PCA Train using SVM Classify using SVM Show Recognized alphabet Computer vision Cascade Object in this experiment, dissimilarities hand gesture representations in shape are also found. These are alphabets 'က', 'ခ', 'ည', 'ဠ', 'ဒ', 'ဗ', 'ဘ', 'ဝ', 'သ', 'ဟ' and 'ဋ'.



The recognized char is : O

Fig.6. Result

No.	Myanmar Alphabet	No. of No. of Testing	No. Correct Time	Success Rate
1	က	20	18	90%
2	ခ	20	16	80%
3	ဂ	20	16	80%
4	ဃ	20	18	90%
5	င	20	18	90%
6	စ	20	17	85%
7	ဆ	20	17	85%
8	ည	20	18	90%
9	ဋ	20	20	100%
10	ဌ	20	17	85%
11	ဍ	20	20	100%
12	ဎ	20	18	90%
13	တ	20	18	90%
14	ထ	20	17	85%
15	ဒ	20	17	85%
16	ဓ	20	18	90%
17	န	20	18	90%
18	ပ	20	18	90%
19	ဖ	20	17	85%
20	ဗ	20	20	100%
21	ဘ	20	20	100%
22	မ	20	16	80%
23	ယ	20	18	90%
24	ရ	20	16	80%
25	လ	20	17	85%
26	ဝ	20	18	90%
27	သ	20	18	90%
28	ဟ	20	18	90%
29	ဠ	20	20	100%
30	အ	20	17	85%

Table1. Recognition Accuracy

VI. CONCLUSIONS

The proposed system can be used for real time recognition. This system can recognize 30 static Myanmar gestures except three dynamic hand gestures (,and). The proposed system can convert the human hand gestures into text with high accuracy and least time.

This research work can be extended to recognize the rotation and distance invariant MSL alphabets gestures, numbers gestures and other complex gestures in different background, location, lighting conditions in real time environment. This research work can also be extended to recognize Myanmar words and sentences.

REFERENCES

[1] L. Pigou et al. "Sign Language Recognition Using Convolutional Neural Networks". European Conference on Computer Vision 6-12 September 2014.

[2] Singha, J. and Das, K. "Hand Gesture Recognition Based on Karhunen-Loeve Transform", Mobile and Embedded Technology International Conference (MECON), January 17-18, 2013, India. 365-371.

[3] Liang, R. H. and Ouhyoung, M.: A Real-time Continuous Gesture Recognition System for Sign Language, The 3rd International Conference on Automatic Face and Gesture Recognition, (1998).

[4] Christopher, L. and Xu, Y.: Online Interactive Learning of Gestures for Human Robot Interfaces, Carnegie Mellon University, The Robotics Institute, Pittsburgh, Pennsylvania, USA, (1996).

[5] H. Suk et al. Hand gesture recognition based on dynamic Bayesian network framework. Patter Recognition 43 (9); 3059-3072, 2010.

[6] Kadous, M. W.: Machine Recognition of Australian Signs Using Power Gloves: towards Large-lexicon Recognition of Sign Language, Workshop on the Integration of Gestures in Language and Speech, (1996) 165-174.

[7] Akyol, S. and Canzler, U.: *An Information Terminal Using Vision Based Sign Language Recognition*, ITEA Workshop on Virtual Home Environments, VHE Middleware Consortium, (2002).

[8] Grobel, K. and Assan, M.: Isolated Sign Language Recognition Using Hidden Markov Models, The IEEE International Conference on System, Man and Cybernetics, (1996) 162-167.

[9] Assaleh, K., Shanableh, T. and Zourob, M.: Low Complexity Classification System for Glove-based Arabic Sign Language Recognition, In Neural Information Processing, (2012) 262-268.

[10] Grimes, G. J.: Digital Data Entry Glove Interface Device, U.S. Patent and Trademark Office, (1983).

[11] Hsu, R. L., Motalleb, M. A. and Jain, A. K.: Face Detection in Colour Images, IEEE Transaction on Pattern Analysis and Machine Intelligence, 24(5) (2002) 696-706.

[12] Dunteman, G. H. “Principal Component Analysis,” Sage publications, (1989).