

Analysis and Performance Comparison of the Feature Vectors in Recognition of Malaysian Sign Language

Yona Falinie A. Gaus, Farrah Wong, Renee Chin, Rosalyn R. Porle, and Ali Chekima

Abstract—In this paper, extraction of suitable feature vector as well as the analysis and performance comparison of the feature vectors using Hidden Markov Model (HMM) are presented. Extracting suitable features comprising of centroids, hand distance and hand orientations is a necessary step to represent isolated Malaysian Sign Language (MSL) to enable detection of right and left hand blobs. Then, each feature vector is modeled using HMM and trained to produce its gesture class. By increasing the number of states starting from 3 until 57 states, each feature vector is trained using HMM so that in the recognition phase it could give the maximum probability among all the other HMMs for a specific word. The system performance of the recognition step was evaluated for each feature vector from the trained model, starting from separated feature vector, followed by combined feature vectors and finally, the union feature vectors. In the experiments, we have tested our system to recognize 112 MSL and found that the union feature vector gives the best recognition rate, which is 83%.

Index Terms—Feature vector, gesture path, hand distance and orientation, hidden markov model.

I. INTRODUCTION

Gestures are expressive and meaningful body motions used in daily life mainly for deaf and hearing-impaired person. Therefore, many researchers have aimed to provide natural ways for human-computer interaction (HCI) through automatic gesture recognition. When the sign language is performed, special devices will capture the rotation, movement information and also the degree of freedom of the hands [1], [2]. The system in [3] however used vision based method in which, hand gesture trajectory is identified through skin colors with blobs computation and then, the location of each hand is tracked using a Kalman Filter. In hand gesture recognition technique, Artificial Neural Network (ANN) and

Hidden Markov Model (HMM) is the most widely known technique for recognition purpose. Mekala *et al.*, [4] used combinational neural network model to develop the recognition of gestures using the features computed from the video stream to detect all the alphabets from A to Z where, 100% recognition rate has been achieved.

Elmezain and Al-Hamadi [5] developed a system to recognize the alphabets letter from A to Z and Arabic number from 0 to 9 in real-time by using stereo color image sequences using HMM Left-Right-Banded (LRB) topology.

Manuscript received October 31, 2012; revised November 28, 2012. This work was supported by MOHE Fundamentals Research Grant (FRGS0026-TK-1/2006) obtained through UMS.

The authors are with the School of Engineering and Information Technology, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia (e-mail: yonafalinie@gmail.com; farrah@ums.edu.my).

II. FEATURE VECTORS EXTRACTION

Given a set of video frames, it is required to detect and track the hand gesture movement over time. However, the set of video frames need to undergo skin segmentation before the hand gesture trajectory can be tracked. YCbCr color space and Gaussian Mixture Model (GMM) are used to cluster the skin region [6]. Fig. 1 shows the skin segmented image from the video frames, which consist of right hand (RH) blob (dark blue marker), left hand (LH) blob (green marker) and head (H) blob (cyan marker).

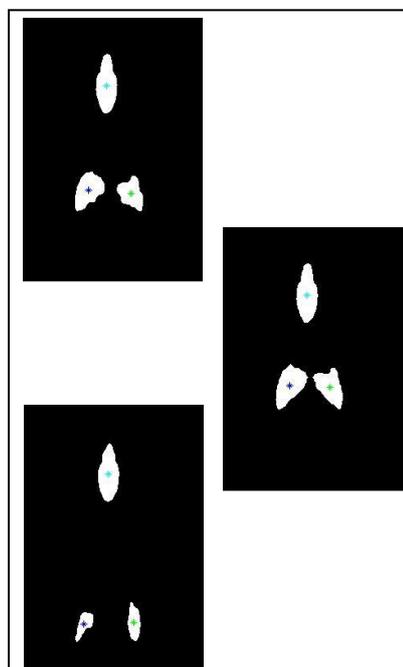


Fig. 1. Skin segmented frames in binary.

A. Centroids

In MSL video sequence, centroids (C_x , C_y) are suitable to define a single point on the hand and the trajectory can be obtained by joining the coordinates of this reference point in every binary frame in the sequence to form gesture path. Using equations (1), (2) and (3), the centroids of each frame are collected. Fig. 2 shows the collected centroids after the completion of centroid extraction from a single MSL gesture.

$$C_x = \frac{\int x \partial A}{A} \quad (1)$$

$$C_y = \frac{\int y \partial A}{A} \quad (2)$$

$$A = \int f(x) \partial x \quad (3)$$

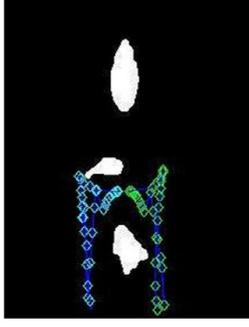


Fig. 2. Collected centroids.

B. Hand Distance

The distance between RH and LH blob with respect to H blob can be calculated using the Pythagoras theorem in equations (4) and (5). Fig. 3 shows the illustration of hand distance, which is measured in pixels.

$$\text{distance RH} = \sqrt{(x_{H_{blob}} - x_{R_{blob}})^2 + (y_{H_{blob}} - y_{R_{blob}})^2} \quad (4)$$

$$\text{distance LH} = \sqrt{(x_{H_{blob}} - x_{L_{blob}})^2 + (y_{H_{blob}} - y_{L_{blob}})^2} \quad (5)$$

C. Hand Blob Orientation

The angle of the hand orientation is measured in the range of -90° to 90° from the horizontal axis by enclosing the blob in an ellipse. The angle between the major axis of the ellipse and the horizontal axis gives the hand tilting angle, representing the orientations of the hand as shown in Fig. 4. $\theta_{handblob}$ in Fig. 4 indicates the angle of enclosing blobs in an ellipse.

III. HAND GESTURE MODELING AND RECOGNITION

HMM is a mathematical model of stochastic process where these processes generate a random sequence of outcomes according to certain probabilities [7]. The feature vectors are modeled using Left-Right-Banded (LRB) topology as shown in Fig. 5. The same model is also used to train the feature vectors for the left hand.

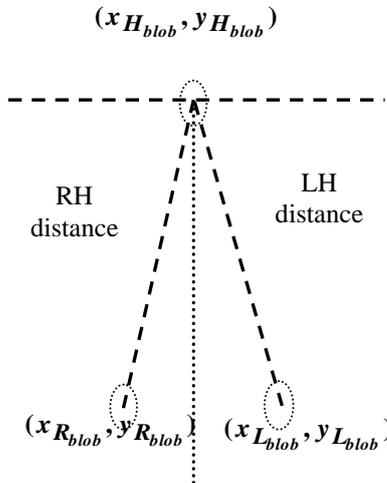


Fig. 3. Distance between head and hand blobs, measured in pixel.

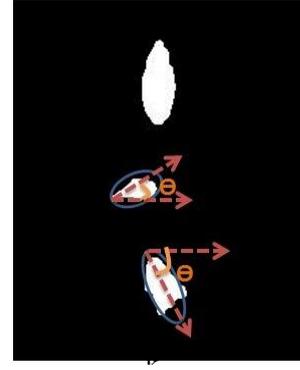


Fig. 4. Hand tilting posture

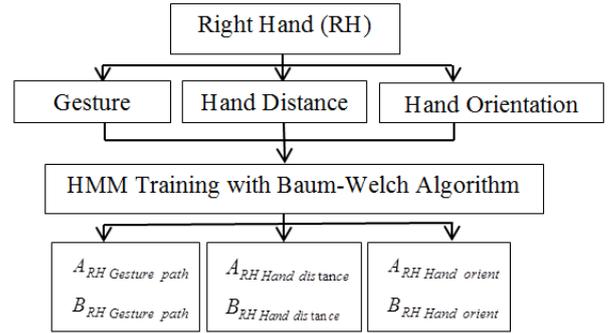


Fig. 5. Separate data training with HMM.

Finally, there are six separate training data (3 from each hand) which are independent of each other. Then, each value will be summed up to represent the output with a single value as given in the equation in (6) for comparison with the results using test data:

$$\begin{aligned} \log[P(O|\lambda_{gesture\ MSL})] &= \log P(O|\lambda_{gesture\ path}) \\ &+ \log P(O|\lambda_{hand\ dist}) \\ &+ \log P(O|\lambda_{hand\ orient}) \end{aligned} \quad (6)$$

IV. EXPERIMENTAL RESULTS

There are 112 selected MSL signs. Each sign has been performed 6 times by different signer, capturing a single hand or both hands moving in different directions with constant or time-varying hand shape. So, in total, there are 672 video sequence of MSL, of which, 560 video sequence are used for training and the remaining 112 videos are used for testing. The experiments for training data begin by increasing the number of states starting from 3 states until 57 states.

In order to check the system performance, the test data is taken for separate feature vector, combined feature vector and union feature vector as shown in Fig. 6 while recognition results is shown in Table I. From Fig. 6, it can be seen that the recognition rate greatly improved when the feature vectors are combined and union compared to the separated feature vector. It is because, while the gesture recognition module needed to match the feature vector against the database of reference gestures, there are chances of the multiple gestures model in HMM modeling that may share the same component of feature vector, which explained why

the separated feature vector gives low accuracy results.

However, it is rare for the gesture to share the same component feature vector, in two or three components. It can be seen that the highest number of recognition rate is largely affected by the number of states. While the effect of increasing the number of states in the HMM modeling on MSL gesture and recognition rates is examined, it is found that, increasing the number of states is in fact increasing the degree of freedom allowed in modeling the MSL gesture training data. Increasing the degree of freedom is an advantage for MSL gesture that has more movement in respective gesture such as “selamat datang”, “selamat pagi”, “selamat petang” and “selamat tengahari”. It is because this respective gesture has higher transition movement, meaning that the hand is making a lot of movement at one time, however, certain gesture do not extent for long duration and only represented by lesser hand movements, which can be modeled by lesser number of states too. If higher number of

states is applied for gesture that has lesser hand movement, it only makes computation time higher while adding redundant data that does not contribute to the recognition rate.

V. CONCLUSIONS

This paper presented a method to recognize the MSL gesture from video sequence by selecting suitable feature vectors that represent the MSL. There are three selected feature vectors that are centroids, hand distance and hand orientations. In the testing phase, it has been conducted by separated feature vector, combined feature vector and union feature vector. Our results show that, union feature vector gives the highest recognition rate that is 83.1%.

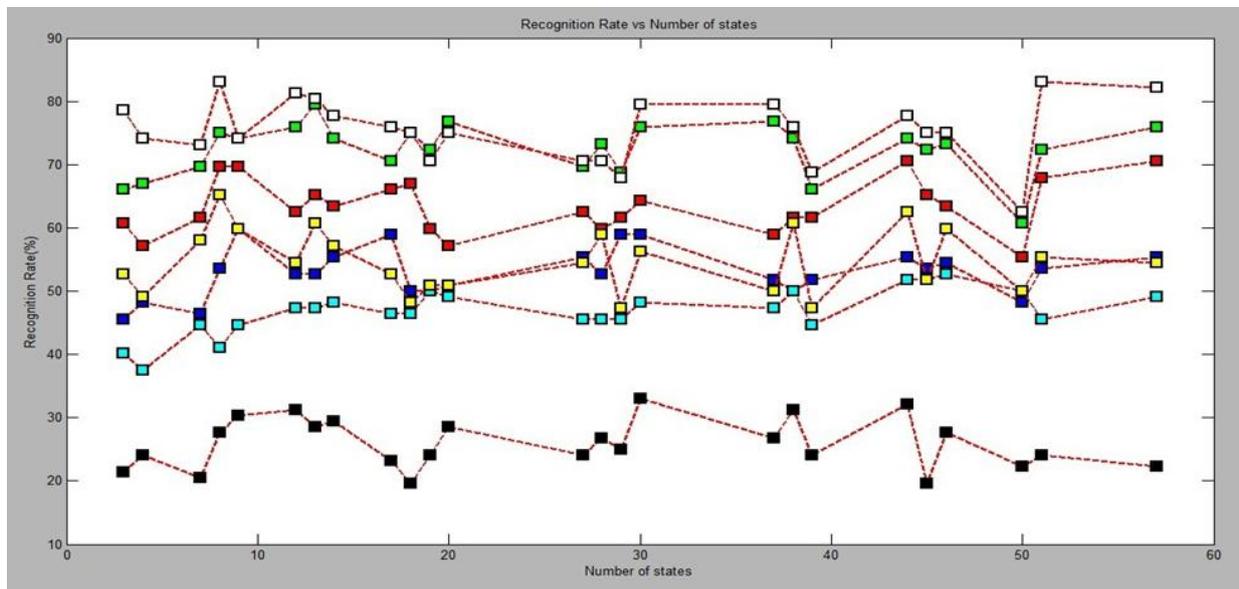


Fig. 6. Recognition rate versus number of state. Color is as indicated in Table I.

TABLE I: RECOGNITION RESULTS OF MSL GESTURES ACCORDING TO FEATURE TYPE WITH AVERAGE RECOGNITION RATE AND THE BEST NUMBER OF STATES

Feature Type	Feature vector	Average Recognition Rate (%)	Number of states	Test Data Results	
				Recognized Data	Recognition Rate (%)
Separated	p , ■	25.92%	38	35	31.25
	q , ■	52.96%	9	67	59.82
	r , ■	46.82%	44, 45	58	51.78
Combined	$(p \cap q)$, ■	63.32%	57	79	70.53
	$(p \cap r)$, ■	54.75%	8	73	65.18
	$(q \cap r)$, ■	72.36%	13	89	79.46
Union	$(p \cap q \cap r)$, ■	75.46%	8, 26, 51	93	83.11

REFERENCES

[1] X. Chen, X. Zhang, Z. Y. Zhao, J. H. Yang, V. Lantz, and K. Q. Wang, “Hand Gesture Recognition Research Based on Surface EMG Sensors and 2D-Accelerometers,” in *Proc. 11th IEEE International Symposium on Wearable Computers*, 2007, pp. 1-4.

[2] H. Brashear, T. Starner, P. Lukowicz, and H. Junker, “Using Multiple Sensors for Mobile Sign Language Recognition,” in *Proc of the 7th IEEE International Symposium on Wearable Computer*, 2003, pp. 45-52.

[3] K. Imagawa, S. Lu, L. Shan, and S. Igi, “Color-Based Hands Tracking System for Sign Language Recognition,” *Biomedical Engineering*, 1998.

[4] P. Mekala, Y. Gao, F. Jeffrey, and A. Davari, “Real-Time Sign Language Recognition Based on Neural Network Architecture,” in *Proc. Joint IEEE International Conference on Industrial Technology &*

43rd Southeastern Symposium on System Theory (SSST'11), Auburn, AL, 2011, pp. 197-201.

- [5] M. Elmezain, A. Al-Hamadi, and B. Michaelis, "Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences," *Gesture*, vol. 16, no. 1, pp. 65-72, 2008.
- [6] Y. Falinie, A. Gaus, and F. Wong, "Hidden Markov Model - based Gesture Recognition with Overlapping Hand- Head / Hand-Hand Estimated using Kalman Filter," in *Proceedings of the ISMS 2012*, Kota Kinabalu, Sabah, 2012, pp. 262-267.
- [7] M. Elmezain and A. Al-Hamadi, "Gesture Recognition for Alphabets from Hand Motion Trajectory Using Hidden Markov Models," in *IEEE International Symposium on Signal Processing and Information Technology*, 2007.



Renee Chin graduated with her Ph.D. degree from the University of Manchester in 2010. She is currently a Senior Lecturer attached to Universiti Malaysia Sabah (UMS). Her research interest includes instrumentation and signal processing.



Rosalyn R. Porle received her B.E. degree (honors) in electronics and telecommunication engineering from Universiti Malaysia Sarawak (UNIMAS) in 2001, M.Sc. degree in electrical and electronics engineering from Universiti Malaysia Sabah (UMS) in 2005 and the Ph.D degree in Computer Engineering from Universiti Malaysia Sabah (UMS) in 2011. Currently she is a senior lecturer in the School of Engineering and

Information Technology, Universiti Malaysia Sabah, Malaysia. Her current research interests include image processing and artificial intelligence.



Yona Falinie was born in Kota Belud, Sabah in 1986. She received her Bachelor Degree in Electrical and Electronic Engineering in 2009 from Universiti Malaysia Sabah (UMS). She is currently pursuing her Master Degree in UMS. She has joined IEEE, USA as a student member since 2006. Her current work is related to image processing and artificial intelligence.



Farrah Wong graduated with her Ph.D. degree from UMS in 2004. She has been elevated to Senior Member, IEEE, USA recently. She is currently a Senior Lecturer attached to Universiti Malaysia Sabah (UMS). Her research interest includes machine vision and intelligent systems.



Ali Chekima received his BEngg in Electronics from Ecole Nationale Polytechnique of Algiers in 1976 and his MSc and PhD both in Electrical Engineering from Rensselaer Polytechnic Institute Troy, New York, in 1979 and 1984 respectively. He joined the Electronics Department at the Ecole Nationale Polytechnique, Algiers in 1984, where he was Chairman of the Scientific Committee of the Department as well as in charge of the Postgraduate Program while teaching at both graduate and undergraduate levels.. He was member of several scientific committees at the national level. He has been working as an Associate Professor at the School of Engineering and Information Technology at Universiti Malaysia Sabah since October 1996. Currently, he is Professor and Head of the Computer Engineering Program. His research interests include Source Coding, Antennas, Signal Processing, Pattern Recognition, Medical Imaging, Biometrics, Data Compression, Artificial Intelligence and Data Mining. He has published more than 140 papers in refereed journals, conferences proceedings, book chapters and research reports.