International Journal of Advanced Computer Research
ISSN (Print): 2249-7277 ISSN (Online): 2277-7970
Volume-5 Issue-19 June-2015

Key Frame Selection for One-Two Hand Gesture Recognition with HMM

Ketki P. Kshirsagar*
Department of Electronics and Telecommunication Engineering, Working in Vishwakarma Institute of Information Technology College of Engineering, Pune, India

Received: 21-May-2015; Revised: 16-June-2015; Accepted: 19-June-2015

Abstract

The sign language recognition is the most popular research area involving computer vision, pattern recognition and image processing. It enhances communication capabilities of the mute person. In this paper, I present an object based key frame selection. Forward Algorithm is used for shape similarity for one and two handed gesture recognition. That recognition is with feature and without feature using HMM method. I proposed use to the hidden markov model with key frame selection facility and gesture trajectory features for one and two hand gesture recognition. Experimental results demonstrate the effectiveness of my proposed scheme for recognizing One Handed American Sign Language and Two Handed British Sign Language.

Keywords

Sign language, Hidden markov model and Trajectory feature for one handed and two handed sign.

1. Introduction

Sign language is not universal; it changes from country to country. Every country has its unique interpreter. Recognition of sign language is to provide most important opportunity for deaf community. Sign language provides an opportunity for a mute person to communicate with normal or mute person without any interpreter. The work on sign language recognition is reported by Starner and Pentland [1], [2], who developed a glove-environment system capable of recognizing a subset of the American Sign Language (ASL). Liang and Ouhyoung [3] used the hidden markov model (HMM) approach for recognition of continuous Taiwanese Sign Language with a vocabulary of 250 signs. Yang and Ahuja [4] used Time-Delay Neural Networks (TDNN) to classify motion patterns of ASL. Bhuyan, et.al. [5] developed recognition of Indian sign language with a vocabulary of 48 signs. For segmentation, Meier and Nagan [6] had developed hausdorff distance algorithm. Though different researchers proposed different methods for sign language recognition, none of these methods are successful to address all the problems encountered in sign language recognition.

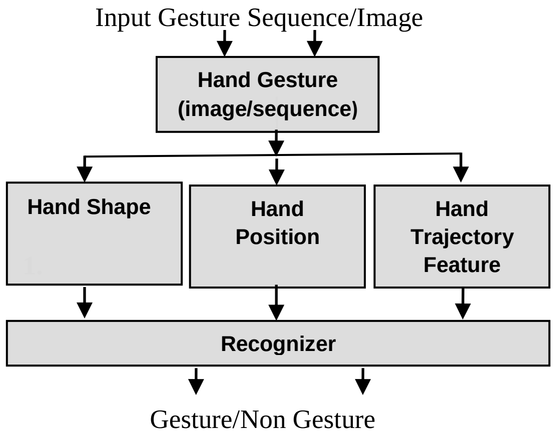


Figure 1: Basic recognition system

In my system, recognition depends upon hand shape, hand position, static trajectory features and dynamic trajectory features. The basic recognition system diagram is as shown figure 1. These are main attributes of hand gesture recognition. In this system, hidden markov model (HMM) is used for one as well as two hand gesture recognition. HMM (with trajectory feature and without feature) includes key frame selection and forward algorithm for observed probability of the given sequence/image, pixel-to-pixel distance measured by lowest weight distance.

*Author for correspondence

The main contributions in this paper are summarized as follows:

Firstly, we propose novel technique for one and two hand gesture recognition to select key frames with HMM (with trajectory features and without feature) method. The advantage of proposed method is that shape and pixel distance measurement for key frames are required instead of all frames of video sequences. The key frame based gesture representation is equally useful for quick gesture recognition and coding of video frames in compressed domain. Detection co-articulation or non-gesturing phase is possible by extracting the key frames in video sequence. Secondly to validate our result we have used two set of databases for one and two hand recognition: first database consist of static alphabet (A to Z) signs, one handed of American Sign Language and two handed of British Sign Language and second database consist of dynamic alphabet sequences one handed of American Sign Language and two handed of British Sign Language. The result of proposed method found outstanding with compared to other existing methods. Rest of the paper is organized as follows. In section 2 I describe proposed hand gesture recognition system. Experimental results are given in section 3. Conclusions are given in section 4.

2. Proposed Gesture Recognition System

The proposed hand gesture recognition system is shown in figure 2.where the input is the video sequence or static image. The first step is converted video sequence into frame format (any size of frame format). Backgrounds in the used video sequences are uniform and non-uniform. In hand gesture recognition first step is video object plane (VOP) generation.

2.1. VOP (Video Object Plane) generation

Inter-frame change detection algorithm is used for extracting the VOP using contour mapping. It is one of most feasible solution to object tracking. For removing the background, skin color segmentation is used [7] as shown in equation (1).

In this process, sequence frames or static image are converted into gray scale. Canny operator is used for luminance edge mapping.

$$\left[\begin{array}{l} (R>95) \quad \& \quad (G>40) \quad \& \quad (B>20) \\ \& \quad (\max\{R,G,B\}-\min\{R,G,B\}>15) \\ \& \quad (|R-G|>15) \quad \& \quad (R>G) \quad \& \quad (R>B) \end{array}\right] \quad (1)$$

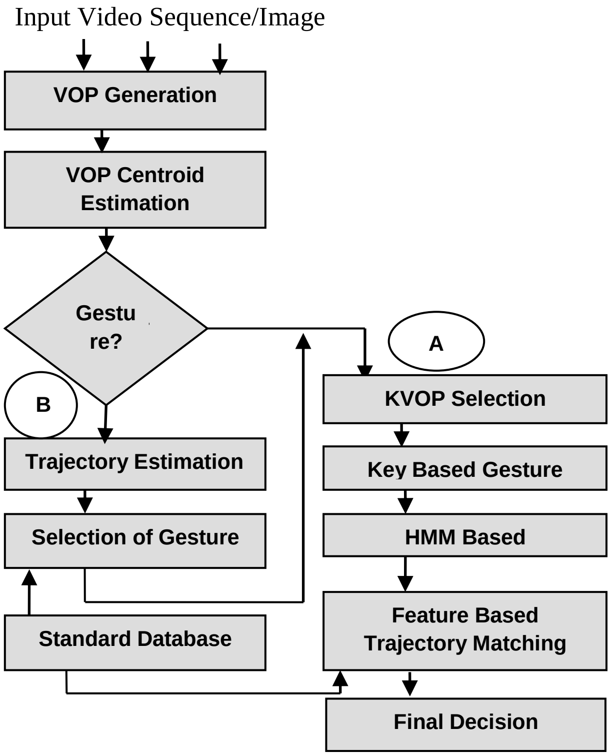


Figure 2: Proposed hand gesture recognition system

A -------- Local motion only.
B -------- Global only or local and global motion.

Difference edges (DE_n) is computed between two successive frames using equation that is the inter-frame change detection algorithm (2)

$$DE_n = \phi(|F_{n-1} - F_n|) \quad (2)$$

Moving change edges (ME_n^{change}) are calculated using difference edges (DE_n) and current frame edges (E_n) using equation (3). Moving still edges (ME_n^{still}) are calculated using moving edges of previous frame (ME_{n-1}) and current frame edges (E_n) using equation (4).

$$ME_n^{change} = \left\{ e \in E_n \,\middle|\, \min_{x \in DE_n} \|e - x\| \le 1 \right\} \quad (3)$$

$$ME_n^{still} = \left\{ e \in E_n \,\middle|\, \min_{x \in ME_{n-1}} \|e - x\| \le 1 \right\} \quad (4)$$

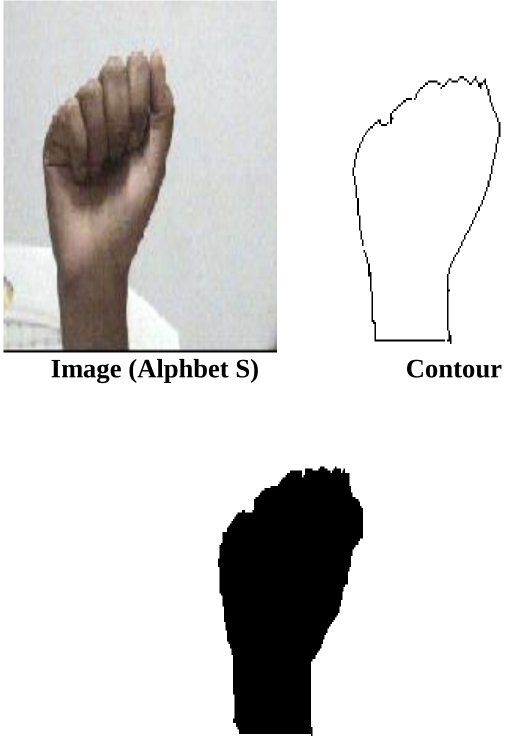


Image (Alphbet S) **Contour**

Binary Alpha Plane

Figure 3: VOP Generated

Using $ME_n^{change}$and ME_n^{still} moving edges (ME_n) are calculated using equation (5).

$$ME_n = ME_n^{change} \cup ME_n^{still} \quad (5)$$

$E_n = \{e_1, e_2, e_3 ... e_n\}$ Current frame edges

Now extract the VOP using contour mapping [8], [9] in figure 3.

Centroid of palm region decides local motion (static) or global motion (dynamic) [5]. There is generally no movement of VOP centroids in case of gestures having only local motions, except for very small movement due to hand trembling. On the other hand, in case of gestures having global motions, VOP centroids will move by large amount from one key VOP to the next. VOP range is shown in figure 4.

As shown in Figure 4, square region indicate local motion range (centroid area). If frame centroid is within the square region it indicates the local motion, and centroid beyond that range indicates global motion.

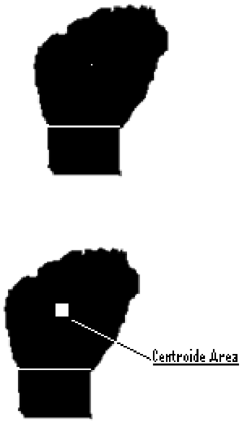


Figure 4: centroid region

In my method, square region is considered around the centroid of the first key VOP and the square space represents the allowable movement of the VOP centroid. The square region is calculated during frames checking and selects the object key frames as well as binary alpha plans that are the segmentation.

2.2. Local motion

Hidden markov model is used for the local motion detection.

2.2.1 Hidden Makov Model

After VOP extraction, binary alpha plane is generated. For the key VOP selection, the first VOP of video sequence is considered as the key VOP. If the canny edge difference between two successive frames is greater than the adaptive threshold, then next frame is selected as next key VOP. The threshold calculated using equation (6).

$$T = (Canny \;\; edge \;\; point\, F_n) \; - \; (Canny \;\; edge \;\; point\, F_{n-1})$$
(6)

F_n - nth frame F_{n-1} - n-1th frame

In HMM, forward algorithm compare two frames (i.e. Test frame sequence T and standard frame S) to finding observed probability of the given sequence (i.e. between State transition matrix and confusion matrix) using equation (7). Here viterbi algorithm is not used because of key frames.

In HMM based recognition observed sequence multiplied are already selected. Matrix and probability calculated by using forward algorithm [10] [11]. If it satisfies criteria, then recognition is successful otherwise it is co-articulation or non-gesture phase.
Sum of all partial probabilities gives probabilities of test given the HMM [12] [13].

$$\alpha_{(t+1)}(j) = \left| \alpha_t(j) \sum_{i=1}^{n} B_{ij} \right| \qquad (7)$$

$$\Pr(\text{matching}) = \left| \sum_{j=1}^{m} \alpha_n(j) \right|$$

Next important step for HMM method in one hand as well as two hands gesture recognition is the selection of suitable features. Selecting good feature is crucial for gesture recognition, because it is totally depend on there shape and motion. For trajectory matching consider both static and dynamic features. Dynamic features used for global motion only. Static features are low level features and dynamic features are high level feature. Static features correspond to shape of hand trajectory and dynamic features correspond to motion of hand trajectory [14]. For gesture recognition both static and dynamic features are equally important. Key trajectory point selection, trajectory length calculation, location feature extraction, orientation feature extraction, velocity and acceleration these are the six features used for recognition.

Key point selection is merging of adjacent approximation interval of the estimated trajectory. These key points best represent the prominent locations of the hand in the gesture trajectory. Trajectory length calculated using equation (8).

$$D = \sum \left\{ \left(S_i - S_{i+1}\right)^2 + \left(T_i - T_{i+1}\right)^2 \right\}^{1/2} \qquad (8)$$

Local feature extraction is the measure of the distance between the center of gravity and the selected key points in a gesture trajectory by equation (9).

$$\hat{s} = \frac{1}{N}\sum_{i=0}^{N} S_i \qquad \hat{t} = \frac{1}{N}\sum_{i=0}^{N} T_i$$

$$L_i = \left| \sqrt{\left(S_i - \hat{s}\right)^2 + \left(T_i - \hat{t}\right)^2} \right| \qquad (9)$$

Orientation feature gives the direction along which the hand traverses in space while making a gesture by equation (10).

$$d_i = \left[s_i - s_{i-1}, t_i - t_{i-1}\right], i = 1, 2, \ldots, N$$

$$\theta_i = \tan^{-1}\left(\frac{t_i - t_{i-1}}{s_i - s_{i-1}}\right), i = 1, 2, \ldots, N \qquad (10)$$

Dynamic feature are the velocity and acceleration. Velocity is based on an important observation that each gesture is made at different speeds calculated by equation (11).

$$v_i = \left\{ \frac{s_{i+1} - s_i}{tm_{i+1} - tm_i}, \frac{t_{i+1} - t_i}{tm_{i+1} - tm_i} \right\}, i = 1, .., N - 1 (11)$$

Acceleration feature which may distinguish co-articulation phase from the meaningful dynamic gesture sequence, as during co-articulation hand moves very quickly by equation (12)

$$\frac{dv_i}{dt} \ldots i = 1, 2, \ldots, N - 1 \qquad (12)$$

2.3 Global motion

Palm region centroid decided the global motion [5]. In global motion most important part is the trajectory estimation. For trajectory estimation first find out the centroid of two successive frames and these are the C_i and C_{i-1}. In global motion, centroid C_i is obtained by translating C_{i-1} respective motion vector in figure 5. The final centroid is taken as the average of these two this nullifed the effect of slight shape changes in successive VOPs [5]. After nullifying effect it behaves as local motion and follows the same path for recognition.

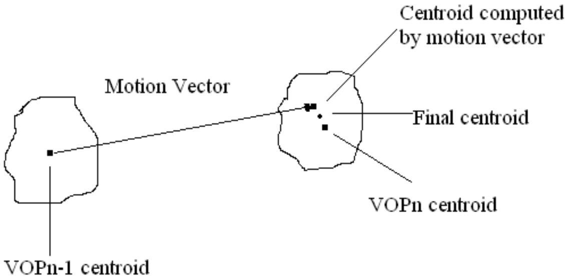


Figure 5: Estimate centroid in VOPs

3. Experimental Result

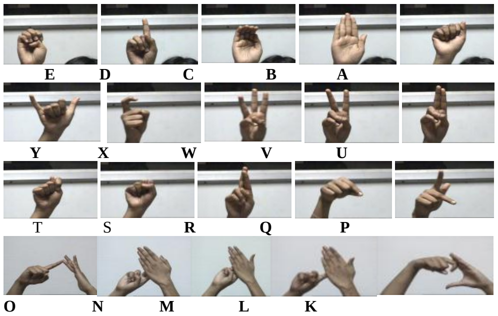
For recognition we have used two set of databases: first database consist of static alphabet (A to Z) signs and second database consist of dynamic alphabet sequences of American Sign Language for one handed and first database consist of static alphabet (A to Z) signs and second database consist of dynamic alphabet sequences of British Sign Language for Two handed. These signs are available in lifeprint fingurespell library [15][16][17]. For recognition of static hand gesture 6500 alphabet signs (Few static alphabet one and two handed signs shown in figure 6) and for dynamic hand gesture 130 sequences are used of the five different persons. All the signs are single handed gestures as well as two handed gestures and video sequences. Table I and Table II show the average recognition efficiency without and with feature of our proposed HMM method.

Table I: Average Gesture Recognition efficiency in % (without feature)

	Method	HMM
One Handed	Static	71.626%
	Dynamic	69.830%
Two Handed	Static	73.58%
	Dynamic	71.79%

Table II: Average Gesture Recognition efficiency (with feature)

	Method	HMM
One Handed	Static	73.53%
	Dynamic	71.672%
Two Handed	Static	75.38%
	Dynamic	72.353%



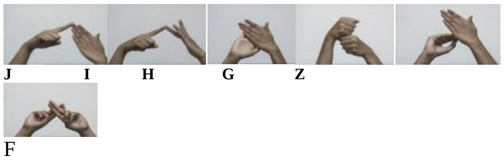


Figure 6: Static alphabet signs (A to Z)

4. Conclusions

The proposed hand gesture recognition system can used different type gesture signs. In this system recognition is done by proposed HMM. Proposed system is suitable for static as well as dynamic one hand and two hand gesture recognition. Advantage of proposed system is the instead of checking all frames in sequence only key frames are checked. Key frame based gesture recognition is more beneficial for fast recognition. And trajectory features improve recognition efficiency.

In future work, we would like to develop complete sign language recognition system [18] by using other body parts i.e., head, arm, facial expression etc. The recognition rate of the proposed system is very much promising for future research in this area.

Acknowledgment

Authors would like to thank AICTE, New Delhi, for providing financial support to carry out this project under RPS sanctioned vide F.No.: 8023/BOR/RPS-67/2006-07, dated 26.02.2007.

References

[1] T. Starner and A. Pentland, "Real-time american sign language recognition from video using hidden markov models", Technical Report, M.I.T Media Laboratory Perceptual Computing Section, Technical Report No.375, 1995.

[2] T. Starner, J. Weaver and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no.12, pp. 1371 -1375, 1998.

[3] Liang, Rung-Huei, and Ming Ouhyoung. "A real-time continuous gesture recognition system for sign language." Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. IEEE, 1998.

[4] M.K. Bhuyan, D. Ghosh and P.K. Bora, “Estimation of 2D motion trajectories from

video object planes and its application to hand gesture recognition", Lecture Notes in Computer Science (Springer-Verlag) Pattern Recognition and Machine Intelligence, vol. 3776, pp. 509-514, 2005.

[5] M.K. Bhuyan, D. Ghoah, and P.K. Bora, "A Framework for Hand Gesture Recognition with Applications to Sign Language", Annual Indian conference, IEEE-2006.

[6] Thomas Meier and King N. Ngan, "Automatic video sequence segmentation using object traking", Speech and image technologies for computing and telecommunication. TECON IEEE-1997.

[7] Hebert Luchetti Ribeiro, Adilson Gonzaga, "Hand image segmentation in video sequence by GMM: a comparative analysis" SIBGRAPI 2006: 357-364.

[8] Kim, Changick, and Jenq-Neng Hwang. "A fast and robust moving object segmentation in video sequences." Image Processing, 1999. ICIP 99. proceedings. 1999 International Conference on. Vol. 2. IEEE, 1999.

[9] Changick Kim and Jenq-Neng Hwang, "Fast and Automatic video object segmentation and tracking for content based application", IEEE transactions on circuits and systems for video technology, VOL.12, NO. 2, Feb 2002.

[10] Nair, Vinod, and James J. Clark. "Automated visual surveillance using hidden markov models." International Conference on Vision Interface. Vol. 93. 2002.

[11] Lawrence Rabiner, "A tutorial on hidden markov model and selected application in speech recognition", Proc. IEEE, VOL. 77, No. 2, Feb 1989.

[12] Lee, Hyeon-Kyu, and Jin H. Kim. "An HMM-based threshold model approach for gesture recognition." Pattern Analysis and Machine Intelligence, IEEE Transactions on 21.10 (1999): 961-973.

[13] Sebastien Marcel, Olivier Bernier, Jean–Emmanuel Viallet and Daniel Collobert, "Hand Gesture Recognition using Input–Output Hidden Markov Models", Automatic Face and Gesture Recognition, 2000, Proceedings of Fourth IEEE International Conference on Source: DBLP, pp1-6.

[14] M.K. Bhuyan, D. Ghosh and P.K. Bora, "Key video object plane selection by MPEG-7 visual shape descriptor for summarization and recognition of hand gestures", Proc. Fourth Indian Conference on Computer Vision, Graphics and Image Processing , pp. 638-643, 2004.

[15] www.lifeprint.com.(Accessed in 2010) .

[16] www.deafblind.com.(Accessed in 2010).

[17] www.british-sign.co.uk/fingerspelling_alphabet.php.

[18] Milan sonka, Vaclav Hlavac, Roger boye "Image processing, Analysis and Machine vision", Thomson-Engineering, 2007 (Book).

Dr. Ketki P. Kshirsagar born in 27th Jan 1983 and she was completed her BE in 2005 from WIT Solapur and MTech 2008 from SGGS institute nanded and Ph.D. completed in 2014 from SGGS institute Nanded. She is working in Vishwakarma Institute of Information Technology College of Engineering, Pune. She was published papers at ICDIP2009 (07-08 March 2009) International Conference on Digital Image Processing Bangkok, Thailand, ARTCOM 2010 (15-16 Sept 2010) International conference on advances in recent technologies in communication & computing, Kottyam, Kerala and 2012 4th International Conference on Electronics Computer Technology (ICECT 2012), April 6- 8, 2012, Kanyakumari, India.

Email: ketki.kshirsagar@viit.ac.in