Gesture Recognition for Alphabet Characters from Fingertip Motion Trajectory Using LRB Hidden Markov Models.

Liakis Ioannis [ioannis.liakis14@imperial.ac.uk]
Research Intern, Imperial College of London
Electrical and Electronic Engineering Department, Computer Vision and Learning Lab.
Project Supervisor: Dr. Tae-Kyun Kim [tk.kim@imperial.ac.uk]

Abstract

This report describes a method to recognize alphabet characters from a detected fingertip motion trajectory, of hand gestures, using HMM (Hidden Markov Models). For this purpose a dataset of recorded gestures was created. Orientation features from the gestures are quantized to generate code-words. HMM using Left-Right-Banded (LRB) topology is used with different number of states ranging from 3 to 10 and different number of observation symbols ranging from 8, 12, 18 . The best performance is achieved with 5 states and 8 observation symbols.

Introduction

Hand gestures are of much importance in Human-Computer Interaction and the Computer Vision field. They are an active area of research with the objective being to bring the performance of human-machine interaction close to human-human interaction. There are many applications of the hand gesture recognition. Sign-language recognition [1], communication in video conference [2], using a finger as a pointer for selecting options from a menu [3]. Over the last few years there have been many publications for hand gesture and motion recognition for both alphabet characters and numerical ones [4] [5] [6] [7] [8].

Hidden Markov Models have become very popular in handwriting recognition and are being used in spatio-temporal pattern recognition. The reason for that is their distinctive ability to learn model parameters from observation sequence through Forward-Backward, Baum-Welch for training and Viterbi algorithm for pattern classification. This report describes a Hidden Markov Model based method to recognize alphabet characters from an already detected fingertip motion trajectory.

There are many free datasets available to download online. The gestures in them are recorded with the subject sitting in front of the camera, which is recording either the motion trajectory of the hand centroid or the motion trajectory of a fingertip. A new dataset was necessary because the dataset on this report was recorded with the subject wearing the camera, so that it would record the fingertip motion from the subject's point of view. The recorded dataset contains fingertip motion trajectories which represent writing gestures. These, are spatio-temporal patterns. Therefore, it is reasonable for differences to occur amongst people. So, a time-independent method is used to define the trajectories, as a sequence of directional angles which are then translated into the observation symbols and characterize a gesture. Each letter is then mapped to a different Hidden Markov Model.

Recognition System Implementation

The system is designed to classify the gesture path generated from a single fingertip motion trajectory by using HMM. Using the same system gesture paths generated from hand motion trajectory can also be recognized. The implementation consists of 3 main stages.

(1) **Preprocessing**: track the fingertip to generate its motion trajectory.
(2) **Feature extraction:** enhance the gesture path. Quantize the orientation. Determine the discrete vector (Observation symbols).
(3) **Classification:** the gesture is classified and recognized using the discrete vector and Left-Right Banded model.

A. Preprocessing

This stage has been already implemented by Guillermo Garcia-Hernando, PhD student in Computer Vision and Learning Lab, Imperial College. For the recordings a Creative camera was used. The algorithm was implemented using OpenCV code in Microsoft Visual Studio 2010 C++. Using the camera's set-up infrared depth system, the algorithm detected the hand of the subject and the upper point of the extended index finger (Fig. 1). The subject was able to start and stop recording the fingertip motion tajectory. Each frame sequence was recorded at approximately 40 frames/second at a resolution of 360x240 pixels. The motion trajectory is saved frame by frame both in pictures and in Cartesian system coordinates. The dataset consists of 10 recorded frame sequences per letter, 260 in total.

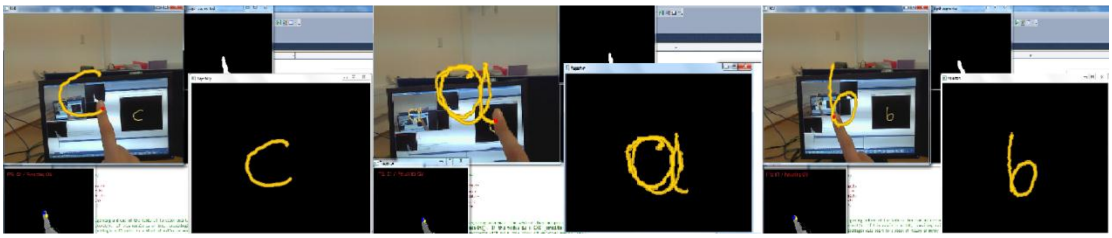

Figure 1: Motion trajectory of detected fingertip

Selecting the best features to describe the gesture path, play a significant role in performance. Three are the basic features: Location, Orientation and Velocity. Orientation is proven to be the most discriminant of the three, followed by velocity and location [5][6][7]. Therefore, in this report, orientation is used as the feature to describe the fingertip gesture trajectory. A gesture trajectory is considered to be the spatio-temporal pattern of the fingertip location L_t.

$$L_t = (x_t , y_t),(t = 1,2,3,\ldots,T) \tag{1}$$

The variable t represents the frame number with T being the last frame and the completion of a gesture. The orientation feature is represented by 'ϑ_t'. It is computed using successive frames of gesture path (Fig 2).

$$\vartheta_t = \arctan\left(\frac{y_{t+1} - y_t}{x_{t+1} - x_t}\right) \tag{2}$$

The extracted feature is further quantized by dividing it, and mapped to different decision regions. At first, only 8 codewords were considered, but in order to compare results to published papers 12 and 18 were considered as well, as in [5][6] (Fig. 2).

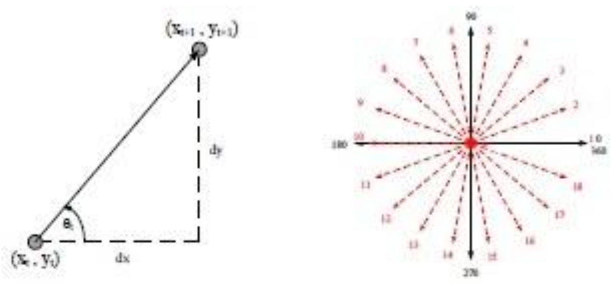

Figure 2. Orientation and its codewords

The feature vectors extracted differ in size, from gesture to gesture, depending on the time of completion for each alphabet, which is affected by the complexity and the shape of the gesture. So, the data were to be aligned in order to be comparable. [9] Presents an algorithm to align data so that the vectors of each alphabet are the same length.

C. Classification

A discrete Hidden Markov system using Left-Right Banded (LRB) topology is used, where each state can lead to itself or the next state only. Hidden Markov Model topologies include Left-Right and Ergodic but in [4] and [10] it is concluded that LRB topology is always better LR and Ergodic. Therefore, in this report LRB topology only is used (Fig 3).

1) ***The Hidden Markov Model***: A Hidden Markov Model is a mathematical model of stochastic processes [11][5]. An HMM is represented by a triple:

$$\lambda = (A, B, \Pi) \tag{3}$$

These parameters represent:

- An N-by-N transition matrix $A = \{a_{ij}\}$, with 'N' being the number of states in every model.

- An N-by-M observation matrix B = $\{b_{im}\}$, with 'M' being the number of observation symbols.
- An initial probability for each state Π_i, i=1,2,...,N.

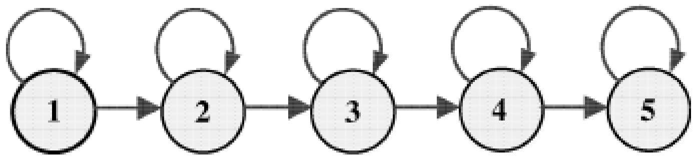

Figure 3: Left-Right Banded Model with 5 states.

2) ***Initializing Parameters:*** About the number of states, the segmented parts of every letter were taken into account. A different number of states per letter and a fixed one were considered with optimum 5 states. The parameters of HMM for the LRB model initialization are the same as in [5]. (Eq.5) describes the matrix A, which depends on the duration time d of states for every alphabet (eq.4).

$$d = \frac{T}{N} \tag{4}$$

$$A = \begin{pmatrix} a_{ii} & 1-a_{ii} & 0 & 0 & 0 \\ 0 & a_{ii} & 1-a_{ii} & 0 & 0 \\ 0 & 0 & a_{ii} & 1-a_{ii} & 0 \\ 0 & 0 & 0 & a_{ii} & 1-a_{ii} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

$$a_{ii} = 1 - \frac{1}{d} \tag{6}$$

The second parameter is also very important. Matrix B is determined by (eq.7), where all the elements of this matrix are initialized with the same value for all the different states.

$$B = \{b_{im}\}\,,\, b_{im} = \frac{1}{M} \tag{7}$$

The third parameter is the initial state distribution vector 'Π' (eq.8) which will ensure that the system initializes from the first state

$$\Pi = (1\;0\;0\;0\;0)^{T} \tag{8}$$

3) Model Training and Evaluation

Separate HMMs are trained for each alphabet using Baum-Welch algorithm. The inputs of this algorithm are a discrete observation vector from feature extraction stage and the initialized parameters. The outputs A,B,Π are an improved guess of the initial parameters, the trained model.

In this report, given test data, the trained models are evaluated using the log-likelihood. To classify a test sequence into one of 26 classes, one per HMM, the log-likelihood of each model and the test sequence is computed; if the first model is the most likely, then the sequence is declared to be of class 1, that is alphabet 'a'. Leave-one-out cross validation is used.

After the sequence classification, the most probable sequence of states for each class is computed using the Viterbi algorithm. Then, each path is compared to the known path from the database and recognized.

Simulation Results

The recognition system was simulated in MATLAB language using Kevin Murphy's HMM toolbox for HMM based system. The simulation was based on 10 recorded frame sequences for every letter, 9 of them were used for training and 1 of them, randomly selected, for testing. The dataset was aligned so that every alphabet's path would be of the same length. The simulation was run for M= **8**, **12** and **18** observation symbols, with 3-10 states, 10 times for each number of states. Results are shown below for fixed number of states. Results for different number of states per alphabet are not shown as they are insignificant to those below.

The successful alphabet classification rate was calculated as the mean of 10 simulations for each different number of states:

$$mean(\ Rc = \left(\frac{Number\ of\ correctly\ classified\ aplhabets}{Total\ number\ of\ aplhabets}\right))$$

The successful recognition rate for each alphabet was calculated as the mean path match rate of 10 simulations for each number of states:

$$mean(\ Rp = \left(\frac{Number\ of\ correctly\ recognized\ states\ in\ path\ sequence}{Total\ path\ length\ of\ each\ aplhabet}\right))$$

Observation symbols M								
8			**12**			**18**		
Number of states	**Rp**	**Rc**	**Number of states**	**Rp**	**Rc**	**Number of states**	**Rp**	**Rc**
3	0.8636	0.8538	**3**	0.8622	0.8731	**3**	0.8612	0.8769
4	0.7949	0.8654	**4**	0.7921	0.8385	**4**	0.7995	0.8615
5	0.7432	0.8962	**5**	0.7402	0.8192	**5**	0.7355	0.8385
6	0.6968	0.8346	**6**	0.7	0.8077	**6**	0.6914	0.8115
7	0.6691	0.8308	**7**	0.6533	0.8115	**7**	0.6507	0.8923
8	0.6206	0.8346	**8**	0.62	0.8346	**8**	0.6164	0.8423
9	0.5858	0.8538	**9**	0.579	0.8423	**9**	0.5721	0.8577
10	0.5624	0.8346	**10**	0.5412	0.8615	**10**	0.5492	0.8154

Best results are achieved with M=8 observation symbols and 5 states. Overall 89,62% of alphabets were correctly classified to the corresponding alphabets. By comparing their path to the known paths from database, there was a 74,32% state sequence match. Really close in alphabet classification was the model with M=18 observation symbols and 7 states, with 89,23% but with poorer path sequence match 65,07% .

Summary and Conclusion

This report describes a gesture recognition method for alphabets, according to the motion trajectory of an index's fingertip. The method has 3 stages. Preprocessing where the trajectory is captured, feature extraction where the orientation is quantized and classification where every fingertip gesture is classified in one of 26 alphabets and compared to said alphabet. Result show that best gesture classification is achieved with M=8 observation symbols and 5 state models with overall 89,62% correct classification and 74,32% path recognition.

References

[1] N. Soontranon, S. Aramvith, T. H. Chalidabhongse, "Improved Face and Hand Tracking for Sing Language Recognition". ITCC 2005, pp. 141-146 Vol. 2, April 2005.

[2] S. Askar, Y Kondratyuk, K. Elazouzi, P. Kauff, o. Schreer, "Vision-Based Skin-Colour Segmentation of Moving Hands for Real-Time Application," 1st European CVMP, pp. 79-85, 2004.

[3] X. Zhu, J. Yang, A. Waibel, "Segmenting Hands of Arbitrary Color," Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 446-453, March 2000.

[4] Nianjun Liu, Brian C. Lovell, Peter J. Kootsookos, and Richard I.A. Davis, "Model Structure Selection & Training Algorithms for a HMM Gesture Recognition System" Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on.

[5] Mahmoud Elmezain, Ayoub Al-Hamadi, Gerald Krell, Sherif El-Etriby, Bernd Michaelis "Gesture Recognition for Alphabets from Hand Motion Trajectory Using Hidden Markov Models", Signal Processing and Information Technology, 2007 IEEE International Symposium on. pp. 1192–1197

[6] Mahmoud Elmezain, Ayoub Al-Hamadi, Jorg Appenrodt, Bernd Michaelis "A Hidden Markov Model-Based Continuous Gesture Recognition System for Hand Motion Trajectory", Pattern Recognition, 2008. ICPR 2008. 19th International Conference on.

[7] Josep Maria Carmona, Joan Climent "A Performance Evaluation of HMM and DTW for Gesture Recognition" Progress in Pattern Recognition, Image Analysis,

Computer Vision, and Applications. Lecture Notes in Computer Science Volume 7441, 2012, pp 236-243 .
[8] Muhammad Zaid Hameed "Initial Project Report Hand Gesture Recognition for Egocentric Depth Video" Project Supervisor: Dr. Tae-Kyun Kim. Department of Electrical & Electronic Engineering Imperial College of Science, Technology & Medicine
[9] Z. Yang, Y. Li, W. Chen, and Y. Zheng, "Dynamic hand gesture recognition using hidden markov models," in Computer Science & Education (ICCSE),2012 7th International Conference on, pp. 360–365, IEEE, 2012.
[10] Appenrodt, J., Elmezain, M., Al-Hamadi, A., Michaelis, B.: "A hidden markov model-basedisolated and meaningful hand gesture recognition". International Journal of Electrical, Computer, and Systems Engineering 3, 156–163 (2009)
[11] Lawrence R. Rabiner. "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". Proceedings of The IEEE - PIEEE 01/1989