Factorial Hidden Markov Models for Gait Recognition

Changhong Chen[1], Jimin Liang[1], Haihong Hu[1], Licheng Jiao[1], and Xin Yang[2]

[1] Life Science Research Center, School of Electronic Engineering, Xidian University
Xi'an, Shaanxi 710071, China
[2] Center for Biometrics and Security Research, Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100080, China
jimleung@mail.xidian.edu.cn

Abstract. Gait recognition is an effective approach for human identification at a distance. During the last decade, the theory of hidden Markov models (HMMs) has been used successfully in the field of gait recognition. However the potentials of some new HMM extensions still need to be exploited. In this paper, a novel alternative gait modeling approach based on Factorial Hidden Markov Models (FHMMs) is proposed. FHMMs are of a multiple layer structure and provide an interesting alternative to combining several features without the need of collapse them into a single augmented feature. We extracted irrelated features for different layers and iteratively trained its parameters through the Expectation Maximization (EM) algorithm and Viterbi algorithm. The exact Forward-Backward algorithm is used in the E-step of EM algorithm. The performances of the proposed FHMM-based gait recognition method are evaluated using the CMU MoBo database and compared with that of HMMs based methods.

Keywords: gait recognition, FHMMs, HMMs, parallel HMMs, frieze, wavelet.

1 Introduction

Hidden Markov models had been the dominant technology in speech recognition since 1980s'. HMMs provide a very useful paradigm to model the dynamics of speech signals. They provide a solid mathematical formulation for the problem of learning HMM parameters from speech observations. Furthermore, efficient and fast algorithms exist for the problem of computing the most likely model given a sequence of observations.

Gait recognition is similar with speech recognition in time-sequential space. Due to the successful application of HMMs to speech recognition, A. Kale, et al, [1, 2] introduced HMMs to gait recognition in recent years and gained inspiring performance. Some other recognition methods [3-5] based on HMMs were proposed one after the other.

There are some possible extensions to the HMMs, such as factorial HMMs (FHMMs) [6], coupled HMMs [7], and so on. FHMMs were first introduced by Ghahramani [6] and attempt to extend HMMs by allowing the modeling of several stochastic random processes loosely coupled. FHMMs are of a multiple layer structure

S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642, pp. 124–133, 2007.

and provide an interesting alternative to combining several features without the need of collapse them into a single augmented feature. In this paper we explore the potential of FHMMs for gait modeling.

This paper is structured as follows. Section II introduces the image preprocessing and feature extraction methods. Section III describes the FHMMs in details and the realization in gait recognition. In section IV, the proposed method is evaluated using the CMU MoBo database [8], and its performances are compared with that of HMMs based methods. Section V concludes the paper.

2 Feature Extraction

2.1 Preprocessing

The preprocessing procedure is very important. The CMU MoBo database [8] offers human silhouettes segmented from the background images. However, the silhouettes are noisy and need to be smoothed.

Firstly, mathematical morphological operations are used to fill the holes and remove some noise.

Secondly, we remove some big noise blocks though filtering, which can't be eliminated by simple morphological operations.

Finally, all the silhouettes are aligned and cropped into the same size. The size can be chosen manually which varies with different databases. For CMU MoBo database, we choose 640*300, which contains most useful information and less noise for most people. An example is showed in Fig. 1.

(a) (b)

Fig. 1. (a) is an example of the original silhouette; (b) is the processed silhouette of (a)

2.2 Feature Extraction

B. Logan [9] pointed out that "there is only an advantage in using the FHMM if the layers model processes with different dynamics; if the features are indeed highly correlated FHMMs do not seem to offer compelling advantages". The choice of features is critical to FHMMs, however, it is really a challenge to choose uncorrelated features from a sequence of gait images.

In this paper, two kinds of different feature extraction methods are employed for different layers of FHMM.

2.2.1 Frieze Feature

The first gait feature representation is a frieze pattern [10]. A two-dimensional pattern that repeats along one dimension is called a frieze pattern in the mathematics and geometry literature. Consider a sequence of binary silhouette images $b(x, y, t)$ indexed spatially by pixel location (x, y) and temporally by time t.

The first frieze pattern is calculated as $F_C(x,t) = \sum_y b(x, y, t)$, where each column (indexed by time t) is the vertical projection (column sum) of silhouette image. The second frieze pattern $F_R(x,t) = \sum_x b(x, y, t)$ can be constructed by stacking row projections. It is considered that F_R contains more information than F_c and some obvious noise can be filtered from F_R as shown in Fig.2. We choose F_R as the feature for the first FHMM layer.

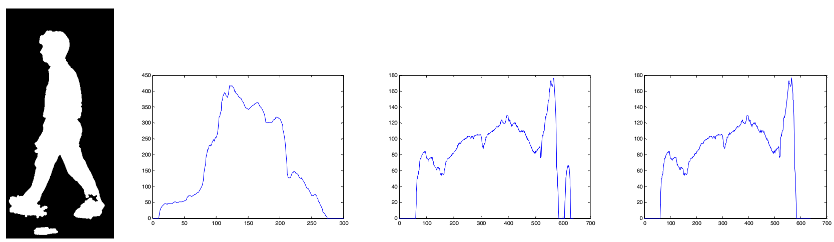

Fig. 2. (a) is a silouette image, its frieze features are F_c (b) and F_R (c). (d) is F_R after filtering noise.

2.2.2 Wavelet Feature

Wavelet transform can be regarded as a temporal-frequency localized analysis method, which has good time resolution in high frequency part and good frequency resolution in low frequency part. It has the property of holding entropy and can change the energy distribution of the image without damaging the information. Wavelet transform acts on the whole image, which can eliminate the global relativity of the image as well as separate the quantization error to the whole image avoiding artifacts.

The wavelet transform suits image processing very much, so we choose the vectors obtained from wavelet transform of the silhouette images as the feature for the second FHMM layer.

3 FHMM-Based Gait Recognition

FHMMs were first described by Ghahramani[6]. They present FHMMs and introduce several methods to efficiently learn their parameters. Our effort, however, is focused on exploiting the application of FHMMs in gait modeling.

3.1 FHMMs Description

The factorial HMM arises by forming a dynamic belief network composed of several layers. Each layer can be considered as an independent HMM. This is shown in Fig. 3. Each layer has independent dynamics but that the observation vector depends upon the current state in each of the layers. This is achieved by allowing the state variable in HMM to be composed of a collection of states. That is, we now have a "meta-state" variable which is composed of states as follows:

$$S_t = S_t^{(1)}, S_t^{(2)}, \cdots S_t^{(M)}, \tag{1}$$

where S_t is the "meta-state" at time t, $S_t^{(m)}$ is the state of the m^{th} layer at time t and M is the number of layers.

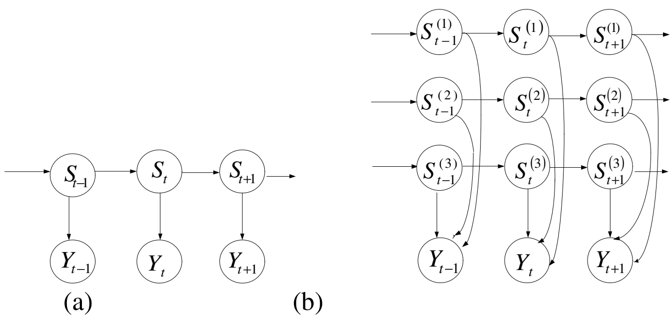

Fig. 3. (a) Dynamic Belief Network representation of a hidden markov model; (b) Dynamic Belief Network representation of a factorial HMM with M=3 underlying Markov chains

It is assumed for simplicity that the number of possible states in each layer is equal. Let K be the number of states in each layer. A system with M layers requires M $K \times K$ transition matrices with zeros representing illegal transitions. It should be noted that this system could still be represented as a regular HMM with a $K^M \times K^M$ transition matrix. It is preferable to use the M $K \times K$ transition matrices over the $K^M \times K^M$ equivalent representation for the computational simplicity.

It is also assumed that each meta-state variable is a priori uncoupled from other state variables:

$$P(S_t \mid S_{t-1}) = \prod_{m=1}^{M} P(S_t^m \mid S_{t-1}^m). \tag{2}$$

As for the probability of the observation given the meta-state, there are two different ways of combining the information from the layers. The first method assumes that the observation is distributed according to a Gaussian distribution with a common covariance and the mean being a linear combination of the state means, which is went by the name of "linear" factorial HMM. The second combination method, the "streamed" method, assumes that $P(Y_t \mid S_t)$ is the product of the distributions of each layer (Y_t is the observation at time t). More details can be found in [9].

3.2 Initialization of Parameters

(1) Number of states K and layers M : Five state numbers are chosen for CMU MOBO database. The number of layers depends on the feature vectors extracted. We extracted two kinds of feature vectors, so the number of layers is two.

(2) The transition matrices: The transition matrices are M $K \times K$ matrices. Each of the initial $K \times K$ matrices is set as a left-to-right HMM, which is only allowed transition from one state to itself and its next state.

(3) Output probability distribution: A gait sequence is always large in size. The large dimension makes it impossible to calculate a common covariance of the observation. So we employ the "streamed" method in 3.1. $P(Y_t \mid S_t)$ is calculated as the product of the distributions of each layer. The models we used are exemplar-based models [2]. The motivation behind using an exemplar based model is that the recognition can be based on the distance measure between the observed feature vector and the exemplars. The distance metric and the exemplars are obviously the key factors to the performance of the algorithm.

Let $Y = \{Y_1, Y_2, \cdots, Y_T\}$ be the sequence of observation vectors, $F^m = \{f_1^m, f_2^m, \cdots, f_T^m\}$ be the feature vectors of the observation vectors in layer m, and T be the length of the sequence. The initial exemplar set is denoted as $S^m = \{s_1{}^m, s_2{}^m, \cdots s_K{}^m\}$. We get the initial exemplar element $s_K{}^m$ by equally dividing observation sequence into K clusters and averaging the feature vectors of each cluster.

We estimate the output probability distribution by an alternative approach based on the distance between the exemplars and the image features. In this way we avoid calculating high-dimensional probability density functions. The output probability distribution of the m^{th} layer is defined as:

$$b_n(f_t^m) = \alpha \delta_n^m e^{-\delta_n^m \times D(f_t^m, S_n^m)}, \tag{3}$$

$$\delta_n^m = \frac{N_n}{\sum_{f_t^m \in e_n^m} D(f_t^m, S_n^m)}, \tag{4}$$

where α is a constant, $D(f_t^m, S_n^m)$ is the inner product distance between the t^{th} feature vector f_t^m and the n^{th} state S_n^m in the m^{th} layer. δ_n^m is defined as equation (4). N_n is the number of frames belonging to the n^{th} cluster, which is constant to all layers. e_n^m represents the n^{th} cluster of the m^{th} layer.

Let β be a constant. The output probability distribution can be represented as:

$$P(Y_t \mid S_t) = \beta \prod_{m=1}^{M} b_n(f_t^m). \tag{5}$$

3.3 Estimation of Parameters

The factorial HMMs we use are exemplar-based. The model parameters are denoted as λ, which include the exemplars in each layer, the transition probabilities between states in each layer and the prior probabilities of each state. The exemplars are initialized as mentioned above and remain unchangeable when estimate other parameters. The transition probabilities and the prior probabilities can be estimated using the Expectation Maximization (EM) algorithm. The algorithm steps can be referred to [6]. The exact Forward-Backward algorithm [6] is used in the E-step. The naive exact algorithm, consisting of translating the factorial HMM into an equivalent HMM with K^m states and using the forward-backward algorithm, has the time complexity of $O(TK^{2M})$. The exact Forward-Backward algorithm has time complexity $O(TMK^{(M+1)})$ because it makes use of the independence of the underlying Markov chains to sum over M $K \times K$ transition matrices. Viterbi algorithm is used to get the most probable path and the likelihood. New exemplars can be obtained through the most probable path, also the new output probability distribution. The whole process is iterated until the likelihood converges to a small threshold.

3.4 Recognition

Firstly, the probe sequence $y = \{y(1), y(2) \cdots y(T)\}$ is preprocessed and extracted features are used as the train sequence.

Then the output probability distribution of the probe sequence can be calculated using the states of the train sequence. We can get the log likelihood P_j that the probe sequence is generated by the FHMM parameters λ_j of the j^{th} person in the train database:

$$P_j = \log(P(y \mid \lambda_j)). \tag{6}$$

The above procedure is repeated for every person in the database. Suppose P_m is the largest one among all P_j's, then we can assign the unknown person to be person m.

A key problem during calculate the log likelihood P_j is how to get the clusters of the probe sequence given the FHMM of the train sequence. We calculate the distance between the features of probe sequence and the exemplars of a train sequence to confirm the clusters. The clusters of the same probe sequence vary with different train sequences.

4 Experiment Results

We use CMU MoBo database [8] to evaluate the proposed method. Fronto-parallel sequences are adopted and the image size is preprocessed to be 640×300. Besides the experiment on the proposed method, other three comparative experiments are conducted. When using only one of the two features, the one layer FHMMs deteriorates to standard HMMs. We give the experiment results of the two HMMs of the two features separately. As showed in Fig. 4, we also give the results of merging the results of the two HMMs. We call this system 'parallel HMM' as [11]. If the judgments of the two HMMs are same, their results will be the results of the 'parallel HMM'. Otherwise, we sum the corresponding likelihoods of the two HMMs and rearrange them to get the final results. Also, the experimental results are compared with that of [1] and [12].

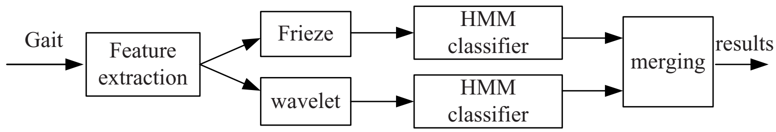

Fig. 4. Parallel HMM

4.1 Same Styles Experiments

The train and probe data sets are of the same motion style. For this type of experiments, we use two cycles to train and two cycles to test.

(a) S vs. S: Training on slow walk of some cycles and testing on slow walk of other cycles.
(b) F vs. F: Training on fast walk of some cycles and testing on fast walk of other cycles.
(c) B vs. B: Training on walk carrying a ball of some cycles and testing on walk carrying a ball of other cycles.
(d) I vs. I: Training on walk in a incline of some cycles and testing on walk in a incline of other cycles.

The results for same style experiments are shown as:

Table 1. The results for same styles experiments

P(%) at rank	HMM[12]		HMM[1]		HMMf		HMMw		pHMM		FHMM	
	1	5	1	5	1	5	1	5	1	5	1	5
S vs. S	100	100	72.0	96.0	100	100	100	100	100	100	100	100
F vs. F	96.0	100	68.0	92.0	88.0	100	100	100	96.0	100	100	100
B vs. B	100	100	91.7	100	95.8	100	100	100	100	100	100	100
I vs. I	95.8	100	---	---	92.0	100	96.0	100	96.0	100	100	100

4.2 Different Styles Experiments

The train and probe data sets are of the different motion styles. For this type of experiments, we use four cycles to train and two cycles to test. The CMC curves for the four experiments of different styles are given in Fig. 5 and the performance comparison with other methods is shown in table 2.

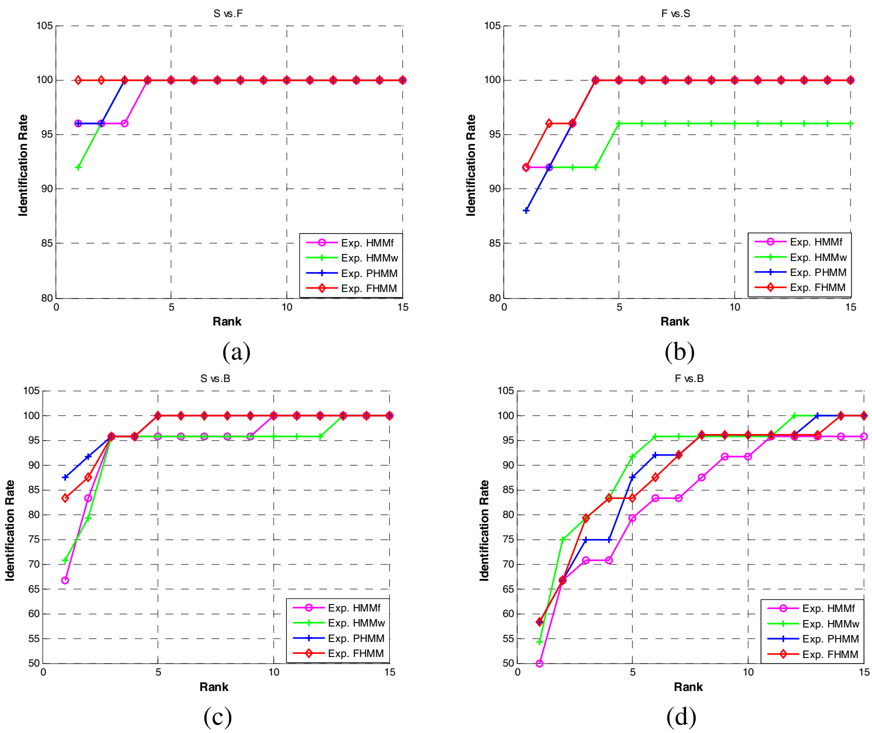

Fig. 5. The cumulative matching characteristics for different styles experiments. Exp.HMMf represents HMM with frieze vectors. Exp.HMMw represents HMM with wavelet transform vectors. Exp. PHMM represents parallel HMM. Exp. FHMM represents factorial HMM. (a) shows the results of S vs. F. (b) shows the results of F vs. S. (c) shows the results of S vs. B. (d) shows the results of F vs. B.

Table 2. The results for different styles experiments

P(%) at rank	HMM[12]		HMM[1]		HMMf		HMMw		pHMM		FHMM	
	1	5	1	5	1	5	1	5	1	5	1	5
S vs. F	---	---	32.0	72.0	96.0	100	92.0	100	96.0	100	100	100
F vs. S	---	---	56.0	80.0	92.0	100	88.0	96.0	88.0	100	92.0	100
S vs. B	52.2	60.9	---	---	66.7	95.8	70.8	95.8	87.5	100	83.3	100
F vs. B	---	---	---	---	50.0	79.2	54.2	91.7	58.3	75.0	58.3	83.3

(a) S vs. F: Training on slow walk and testing on fast walk.
(b) F vs. S: Training on fast walk and testing on slow walk.
(c) S vs. B: Training on slow walk and testing on walking with a ball.
(d) F vs. B: Training on fast walk and testing on walking with a ball.

For same styles experiments, the performance of FHMM-based gait recognition method is excellent, which can reach 100% at rank 1. For different styles experiments, more experiments are done and much better results are obtained than reference [1] and

[12]. For the experiment S vs. F, FHMM-based gait recognition method can reach 100% at rank 1, which is the best result until now. Both experiment S vs. F and F vs. S have gained higher identification rate than experiment S vs. B and F vs. B. When people walk with a ball, their shapes change a lot. Absolutely superiority of FHMM over HMM with a single feature can be seen in all of these experiments. Also the FHMM-based gait recognition method is better that of parallel HMM based method, except the experiment S vs. B.

From the experiment results we can see that the performance of the FHMM-based gait recognition method is superior to that in [1] and [12]. Also its performance is better than the method using frieze feature or wavelet feature individually. Meanwhile, it is a little bit better than parallel HMM. What's more, FHMM is simpler in implement and faster than parallel HMM. The results show that FHMM-based method is effective and improves the performance of HMM.

5 Conclusion

We presented a FHMM-based gait recognition method. The experiment results proved that FHMM is a good extension of HMM. The FHMM framework provides an interesting alternative to combining several features without the need of collapse them into a single augmented feature. FHMM is simpler than parallel HMM in implement. However, the features must be irrelated. It is a challenge problem to extract irrelated but effective features from the same gait sequence. Out future work will concentrate in this area to further improve the performance.

Acknowledgments. This work was partially supported by the Natural Science Foundation of China, Grant Nos. 60402038 and 60303022, the Chair Professors of the Cheung Kong Scholars, and the Program for Cheung Kong Scholars and Innovative Research Team in University (PCSIRT).

References

1. Kale, A., Cuntoor, N., Chellappa, R.: A framework for activity-specific human identification. In: Proc. of the Int. Conf. on Acoustics, Speech and Signal Processing (May 2002)
2. Sundaresan, A., RoyChowdhury, A., Chellappa, R.: A Hidden Markov Model Based Framework for Recognition of Humans from Gait Sequences. In: Proceedings of IEEE International Conference on Image Processing. IEEE Computer Society Press, Los Alamitos (2003)
3. Liu, Z., Malave, L., Sarkar, S.: Studies on Silhouette Quality and Gait Recognition. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04). IEEE Computer Society Press, Los Alamitos (2004)
4. Iwamoto, K., Sonobe, K., Komatsu, N.: A Gait Recognition Method using HMM. In: SICE Annual Conference in Fukui, Japan (2003)
5. Chen, C., Liang, J., Zhao, H., Hu, H.: Gait Recognition Using Hidden Markov Model. In: Jiao, L., Wang, L., Gao, X., Liu, J., Wu, F. (eds.) ICNC 2006. LNCS, vol. 4221, pp. 399–407. Springer, Heidelberg (2006)

6. Ghahramani, Z., Jordan, M.: Factorial Hidden Markov Models. Computational Cognitive Science Technical Report 9502 (Revised) (July 1996)
7. Brand, M.: Coupled hidden Markov models for modeling interacting processes. MIT Media Lab Perceptual Computing/Learning and Common Sense Techincal Report 405 (Revised) (June 1997)
8. Gross, R., Shi, J.: The Cmu Motion of Body (mobo) Database. Technical report, Robotics Institute (2001)
9. Logan, B., Moreno, J.: Factorial Hidden Markov Models for Speech Recognition: Preliminary Experiments. Cambrige Research Laboratory Technical Research Series (September 1997)
10. Liu, Y., Collins, T., Tsin, Y.: Gait Sequence Analysis using Frieze Patterns, CMU-RI-TR-01-38
11. Logan, B., Moreno, P.: Factorial HMMs for Acoustic Modeling, Acoustics Speech and Signal[C]. In: Proceedings of the IEEE International Conference, vol. 2 (S), pp. 813–816. IEEE Computer Society Press, Los Alamitos (1998)
12. Zhang, R., Vogler, C., Metaxas, D.: Human Gait Recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE Computer Society Press, Los Alamitos (2004)