Factorial HMM and Parallel HMM for Gait Recognition

Changhong Chen, Jimin Liang, Heng Zhao, Haihong Hu, and Jie Tian, *Senior Member, IEEE*

***Abstract*—Information fusion offers a promising solution to the development of a high-performance classification system. In this paper, the problem of multiple gait features fusion is explored with the framework of the factorial hidden Markov model (FHMM). The FHMM has a multiple-layer structure and provides an alternative to combine several gait features without concatenating them into a single augmented feature. Besides, the feature concatenation is used to directly concatenate the features and the parallel HMM (PHMM) is introduced as a decision-level fusion scheme, which employs traditional fusion rules to combine the recognition results at decision level. To evaluate the recognition performances, McNemar's test is employed to compare the FHMM feature-level fusion scheme with the feature concatenation and the PHMM decision-level fusion scheme. Statistical numerical experiments are carried out on the Carnegie Mellon University motion of body and the Institute of Automation of the Chinese Academy of Sciences gait databases. The experimental results demonstrate that the FHMM feature-level fusion scheme and the PHMM decision-level fusion scheme outperform feature concatenation. The FHMM feature-level fusion scheme tends to perform better than the PHMM decision-level fusion scheme when only a few gait cycles are available for recognition.**

***Index Terms*—Factorial hidden Markov model (FHMM), gait recognition, information fusion, McNemar's test, parallel HMM (PHMM), performance evaluation.**

I. INTRODUCTION

BIOMETRICS refer to the automatic identification of a person by measuring and analyzing its physiological or behavioral characteristics, such as fingerprints, eye retinas and irises, facial patterns, and gait patterns. Gait recognition is the process of identifying an individual by his/her walking style. In comparison with other biometric characteristics, gait patterns have the advantages of being unobtrusive, difficult to conceal, noninvasive, and effective from a distance. Therefore, gait recognition has attracted a lot of research interests in recent years. Information fusion technology offers a promising solution to the development of a superior classification system. It has been applied to numerous fields and new applications are being explored constantly. For gait recognition problems, information fusion is necessary to be employed under at least three circumstances: 1) when multiple gait features are used for recognition [1]–[5]; 2) when multiple-view gait sequences are employed to develop gait recognition algorithms [6]–[9]; and (3) when gait patterns are integrated with other biometrics [10]–[15].

The problem of multiple gait feature fusion has been addressed by [1]–[5]. Wang *et al.* [1] employed both static and dynamic features for recognition using the nearest exemplar classifier. The features were fused on decision level using different combination rules. Lam *et al.* [2] presented two gait feature representation methods, the motion silhouette contour templates (MSCTs) and static silhouette templates (SSTs), and performed decision-level fusion by summarizing the similarity scores. Bazin *et al.* [3] examined the fusion of a dynamic feature and two static features in a probabilistic framework. They proposed a process for determining the probabilistic match scores using intra- and interclass variance models together with Bayes rule. Han and Bhanu [4] proposed a method to learn statistical gait features from real templates and synthetic templates to address the problem of lacking gallery gait data. A matching score fusion strategy was therefore applied to improve the recognition performance. Veres *et al.* [5] tried to fuse static and dynamic features to overcome the problem when the gallery and probe databases were recorded with a time interval. Generally speaking, superior gait recognition performance was reported when multiple features were employed.

While some research attempted the multiview gait recognition problem by warping the original views to the canonical view [6], [7], others sought information fusion approaches. Wang *et al.* [8] presented a multiview gait recognition method based on fusing the similarity scores of two viewpoints by the sum rule, weighted sum rule, product rule, and Dempster–Shafer rule. Lu and Zhang [9] proposed a multiview fusion recognition approach on the decision level, which combined the results of independent component analysis (ICA) and genetic fuzzy support vector machine (GFSVM) using the product of sum (POS) rule.

Although research conducted in the area of gait recognition has shown the potential of gait-assisted identification, at present, gait is not generally expected to be used as a sole mean of identification of individuals in a large database; instead, it is seen as a potentially valuable component in a multimodal biometric system [16]. See [10]–[15] for approaches to combine gait with other biometrics.

Manuscript received July 15, 2007; revised December 26, 2007 and May 20, 2008. First published December 2, 2008; current version published December 22, 2008. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 60402038, in part by the National Basic Research Program of China (NBRPC) under Grant 2006CB705700, in part by the Chair Professors of the Cheung Kong Scholars, and in part by the Program for Cheung Kong Scholars and Innovative Research Team in University (PCSIRT) under Grant IRT0645. This paper was recommended by Associate Editor L. Zhang.

C. Chen, J. Liang, H. Zhao, and H. Hu are with the Life Science Research Center, School of Electronic Engineering, Xidian University, Xi'an 710071, China.

J. Tian is with the Life Science Research Center, School of Electronic Engineering, Xidian University, Xi'an 710071, China, and also with the Institute of Automation, Chinese Academy of Science, Beijing 100080, China (e-mail: jimleung@mail.xidian.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCC.2008.2001716

In the aforementioned literatures, the approaches of information fusion can be roughly classified into two categories: the decision-level fusion [1]–[11], [13], [15] and the feature-level fusion [12], [14]. In the decision-level fusion system, multiple classifiers work in hierarchical [12], [15] or in parallel [1]–[11], [13], [14]. The outputs of the individual classifiers (subject labels, rank values, or match scores) are combined by some specific fusion rules to produce the final recognition result. Commonly applied fusion rules include majority voting, sum rule, product rule, and so on. Some details of these fusion rules can be found in [3] and [8].

While fusion at the decision level has been extensively studied, feature-level fusion is relatively understudied. Zhou and Bhanu [14] presented a summary of the recent work for the feature-level fusion and pointed out that feature concatenation was the most popular feature-level fusion methodology. Whereas in many recognition applications, the number of available training samples is small, and high-dimensional feature typically results in the well-known curse of dimensionality and the small sample size problem [17].

In this paper, we explore two extended hidden Markov models (HMMs) frameworks, the factorial HMM (FHMM) and the parallel HMM (PHMM), for gait recognition using multiple features. The HMM-based gait recognition methodology is preferable to other techniques since it explicitly takes into consideration not only the similarity between shapes in the test and reference sequences, but also the probabilities with which shapes appear and succeed each other in a walking cycle of a specific subject [16]. The FHMM framework provides an interesting alternative to combine several features without the need to combine them into a single augmented feature vector [18]. It has a multiple-layer structure and combines the information from each layer to obtain the model parameters. We treat FHMM as a feature-level fusion structure because fusion occurs at training process to get model parameters instead of decision level to obtain the recognition results. The PHMM framework, similar with the FHMM, also has a multiple-layer structure. The difference between the PHMM and FHMM is that the PHMM has no interlayer connections and each layer runs independently as an HMM-based classifier. The PHMM acts as a typical decision-level fusion structure. Besides, the feature concatenation is presented as the straightforward feature-level fusion method. In this paper, the PHMM and the feature concatenation are regarded as two traditional fusion methods, whose performances are compared with that of FHMM.

Logan and Moreno [18] 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 FHMM does not seem to offer compelling advantages." Therefore, we chose intuitively the wavelet feature and frieze feature [19] as two distinct features for algorithm validation experiments. Extensive experiments are carried out on the Carnegie Mellon University (CMU) motion of body (MoBo) gait database [20] and the Institute of Automation of the Chinese Academy of Sciences (CASIA) gait database (database A) [21]. The cumulative matching score (CMS) curve is used to evaluate the recognition performance. Besides the CMS curve, McNamara's test is employed to compare the proposed fusion algorithms with the HMM-based approach using the individual feature or the concatenated feature. McNemar's test is a first-order check on the statistical significance of an observed difference in recognition performance. The times of success/failure trials of the compared algorithms are used to calculate the confidence and produce the evaluation result. McNemar's test is generally more comprehensive and reliable than the CMS curve to compare two algorithms, especially when two CMS curves are crossed with each other.

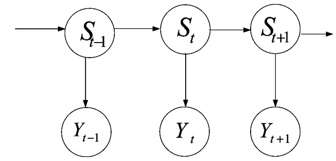


Fig. 1. Hidden Markov model.

In addition to introducing PHMM and FHMM as two kinds of fusion methodology based on HMM, we report another contribution to represent incomplete gait image. The CASIA gait database A was collected outdoor with some incomplete gait silhouettes. The frieze feature does not fit the CASIA gait database A directly. We reprocess each silhouette image in the CASIA gait database A as a frame difference energy image (FDEI). FDEI is represented as the sum of gait energy image (GEI) and the difference image between the current frame and the next frame. The advantage of the FDEI lies in emphasizing the difference of consecutive frames and maximally maintaining the primary shape information.

The remainder of this paper is structured as follows. Section II gives an introduction on FHMM and PHMM. The FHMM-based and PHMM-based silhouette recognition methods are presented in Section III. In Section IV, the proposed methods are evaluated on the CMU MoBo gait database and the CASIA gait database A. Their performances are compared with that of HMM-based methods through CMS curves and McNemar's test. We discuss the results and conclude this paper in Section V.

II. Factoral and Parallel HMM

The structure of HMM is shown in Fig. 1. HMM gets its name from two defining properties. First, it assumes that the observation Y_t at time t is generated by some process whose state S_t is hidden from the observer. Second, it hypothesizes that the state of the hidden process satisfies the Markov property. Given the value of S_{t-1}, the current state S_t is independent from all the states prior to $t-1$.

There are some possible extensions of the HMMs, such as the FHMM [22], the coupled HMM [23], and the PHMM [18]. These extensions have a more complex structure and can achieve better performance if properly used. FHMM and PHMM will be introduced and employed in this paper.

FHMM was first introduced by Ghahramani and Jordan [22], who attempted to extend HMM by allowing the modeling of several stochastic random processes loosely coupled. FHMM uses a more complex state structure to improve the representational

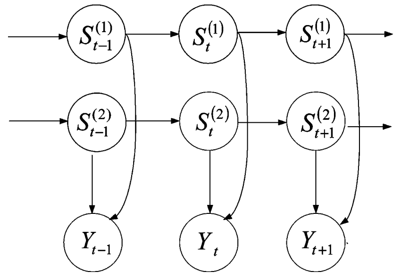


Fig. 2. FHMM with two layers.

capacity of HMM. It has a multiple-layer structure and provides an interesting alternative to combining several features. FHMM was previously applied to speech recognition [18], [24], [25]. Betkowska *et al.* [25] used FHMM for speech recognition with the second layers modeling sudden noise and confirmed it can improve clean speech HMM in noisy conditions. To the best of our knowledge, factorial HMM has not been applied to gait recognition yet. In this paper, we attempt to explore the potential of FHMM for gait modeling as a feature-level fusion method, which combines different features to find an optimal discrimination.

FHMM arises by forming a dynamic belief network composed of several layers. Each layer can be considered as an independent HMM. The structure of FHMM is demonstrated in Fig. 2. Each layer has independent dynamics but the observation vector depends upon the state of all the layers. This is achieved by allowing the state variable in FHMM as a collection of states. We employ a "metastate" variable as the combination of states as follows:

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

where S_t is the "metastate" at time t, $S_t^{(m)}$ indicates the state of the mth layer at time t, and M denotes the number of layers.

For simplicity, it is assumed that each layer of the factorial HMM has the same number of possible states. Let K be the number of states in each layer. An FHMM with M layers requires M $K \times K$ transition matrices with zeros representing illegal transitions. This system could still be represented as a regular HMM with a $K^M \times K^M$ transition matrix. However, for computational simplicity, it is preferable to use the M $K \times K$ transition matrices over the $K^M \times K^M$ equivalent representation. It is also assumed that each metastate variable is *a priori* uncoupled from other state variables, such that the following equation holds:

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

There are two different ways of combining the information from the layers to calculate the probability of the observation. The first method assumes that the observation is distributed according to a Gaussian distribution with a common covariance and the mean from a linear combination of the state means, named as the "linear" FHMM. The second combination method, namely the "streamed" method, assumes that $P(Y_t|S_t)$ is the product of the distributions of each layer (Y_t is the observation at time t). See [22] for more details.

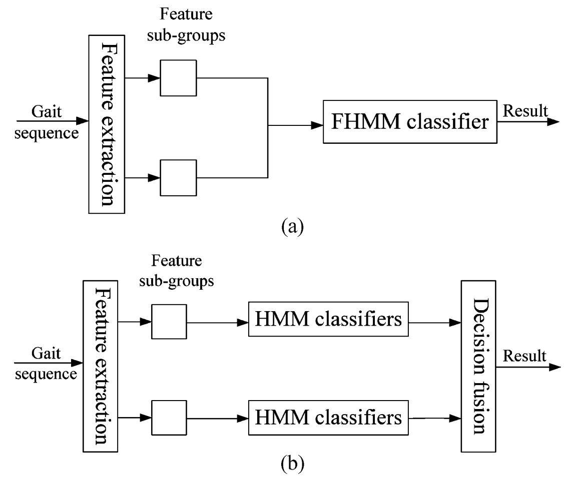


Fig. 3. (a) Fusion structure of FHMM. (b) Fusion structure of PHMM.

FHMM can be considered as a feature-level fusion method and be applied to gait recognition, as shown in Fig. 3(a). In this paper, we use two distinct gait features (as described in Section III) in experiments. The two features are extracted from the same gait sequence but applied to different layers of FHMM. PHMM can be considered as a decision-level fusion structure and the structure of PHMM is displayed in Fig. 3(b). The same two features used in FHMM are trained and tested by two HMM classifiers, respectively. The classification results are fused using the voting rule if the classifiers agree with each other, or using the summation of the matching scores if they are inconsistent.

III. FHMM-Based PHMM-Based Silhouette Recognition

In this section, we describe the FHMM-based and PHMM-based gait silhouette recognition methods in detail. For completeness, we start with an introduction of the preprocessing and feature extraction procedures.

A. Preprocessing and Feature Extraction

The preprocessing procedure is crucial for gait recognition. The CMU MoBo gait database and the CASIA gait database A offer human silhouettes segmented from video sequences. These silhouettes are noisy and need further processing. The preprocessing procedure in this paper has the following steps.

First, mathematical morphological operations are employed for holes remedy and noise elimination.

Second, some big noise blocks, such as the shadow under the feet or some other redundant parts of the silhouettes, are removed by eliminating the connected regions whose areas are less than a given threshold. This step is useful to get rid of the separated noise regions, but may fail when the noise is connected with the main body.

Finally, all the silhouettes are aligned and cropped into the same size. The size can be chosen manually. We choose 640*300 for the CMU MoBo gait database and 120*95 for the CASIA

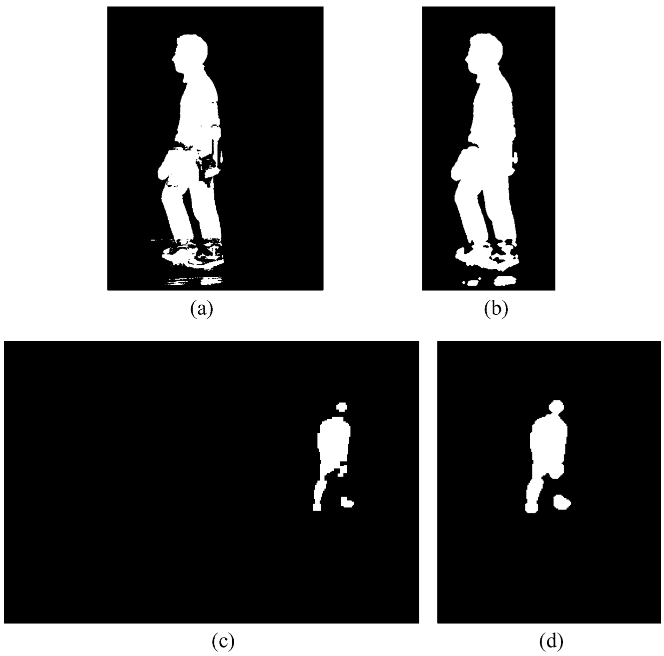


Fig. 4. (a) Example of the original silhouette of the CMU MoBo gait database. (b) Processed silhouette of (a). (c) Example of the original silhouette of the CASIA gait database A. (d) Processed silhouette of (c).

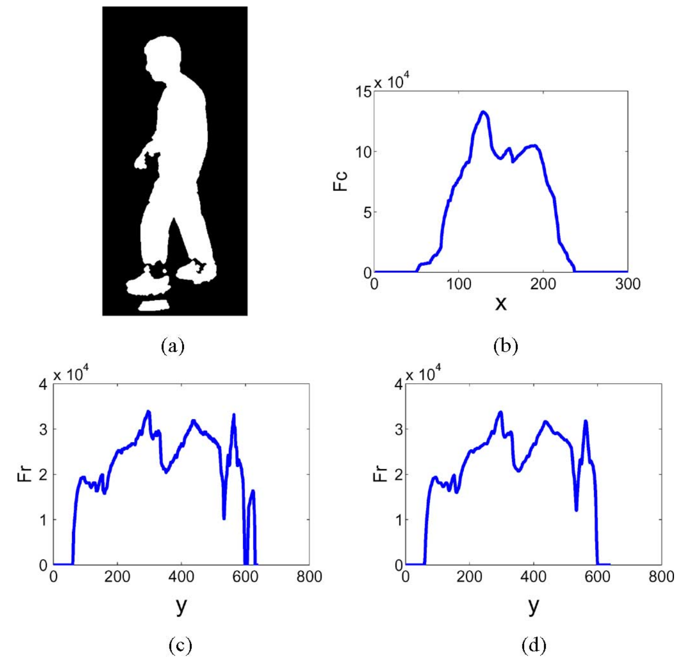


Fig. 5. (a) Preprocessed silhouette image. (b) Frieze feature F_c of (a). (c) Other frieze feature F_r of (a). (d) F_r after filtering noise.

gait database A, which contain most useful information and less noise for most people. Examples are presented in Fig. 4.

After the foregoing processing steps, we proceed to extract features from the silhouette image sequences. Two types of different features, the frieze and wavelet features, are employed. The frieze feature extracts spatial information, while the wavelet feature keeps the low-frequency information.

1) Frieze Feature: The frieze feature pattern was proposed in [19]. A 2-D pattern that repeats along one dimension is called a frieze pattern in mathematics and geometry literature. Consider a binary silhouette image $b(x, y)$ indexed spatially by pixel location (x, y). The first frieze pattern is calculated as $F_c(x) = \sum_y b(x, y)$, which is the vertical projection (column sum) of silhouette image. The second frieze pattern $F_r(y) = \sum_x b(x, y)$ 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 out from F_r, as shown in Fig. 5. We choose F_r as the first gait feature pattern hereafter in the paper.

2) Wavelet Feature: Wavelet transform is regarded as a temporal-frequency localized analysis method, which has both high time resolution in high-frequency part and high-frequency resolution in low-frequency part. It has the property of holding entropy and changing the energy distribution of the image without missing information. Wavelet transform acts on the entire image, which can eliminate the global relativity of the image as well as separate the quantization error to the whole image, thus avoiding artifacts.

The 2-D wavelet transform is applied to the silhouettes using Haar wavelet base. The wavelet coefficients of the approximation subimage hold the most useful information and reduce the dimension of the image. The wavelet coefficients of the approximation subimage are chosen as the second gait feature pattern.

B. FHMM-Based Silhouette Recognition

This section describes the details of the FHMM-based silhouette recognition method. The parameters of FHMM are first initialized and iteratively estimated. The recognition results are then determined by the sorted likelihoods.

1) Parameters Initialization: Number of states K and layers M: As in [26], the average distortion is used to examine the marginal reduction in the distortion as the state number changes. When the average distortion does not decrease rapidly, the corresponding state number is chosen. In this paper, the state number K for the CMU MoBo gait database and the CASIA gait database A are set to be 5 and 7, respectively. The number of layers depends on the features extracted. Since two kinds of features are extracted in this paper, FHMM has two layers, i.e., $M = 2$. *The transition matrices*: The transition matrices are M $K \times K$ matrices. Each of the initial $K \times K$ matrices is initialized as a left-to-right HMM, which only allows transitions from one state to itself and its next state.

The 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 use an exemplar-based model to calculate the distribution. The motivation behind the exemplar-based model is that the recognition can be based on the distance measured between the observed feature vectors and the exemplars. The distance metric and the exemplars are key factors to the performance of an algorithm. The "streamed" method mentioned before in Section II is employed to combine different layers.

Let T be the length of the gait sequence. $\{Y_t, t = 1, \ldots, T\}$ denote the observation vectors of a training cycle, $\{f_t^{(m)}, t = 1, \ldots, T\}$ are the feature vectors, and $\{S_t^{(m)}, t = 1, \ldots, T\}$ is the state set of layer m. The feature vectors are equally divided into K clusters and the exemplar element $S_t^{(m)}$ is initialized as the mean feature vector in layer m of the cluster to which the t-th feature vector $f_t^{(m)}$ belongs.

In order to avoid calculating high-dimensional probability density functions, we estimate the output probability distribution by an alternative approach based on the distance between the exemplars and the feature vectors. The output probability distribution of the layer is defined as

$$P\big(f_t^{(m)} | S_t^{(m)}\big) = \alpha \delta_n^{(m)} \, e^{-\delta_n^{(m)} \times D(f_t^{(m)}, S_t^{(m)})} \tag{3}$$

where α is a parameter less than 1, $D(f_t^{(m)}, S_t^{(m)})$ is the inner product distance between the t-th feature vector $f_t^{(m)}$ and the t-th state in the layer m, $\delta_n^{(m)}$ is a parameter of the cluster nth in the layer m, calculated as:

$$\delta_n^{(m)} = \frac{N_n}{\sum_{t \in C_n} D\big(f_t^{(m)}, S_t^{(m)}\big)} \tag{4}$$

where N_n is the frame number of the nth cluster. C_n contains the frame numbers of the nth cluster.

The output probability distribution is represented by

$$P(Y_t | S_t) \propto \prod_{m=1}^{M} P\big(f_t^{(m)} | S_t^{(m)}\big). \tag{5}$$

2) Parameters Estimation: The FHMM model parameter is denoted as λ, which includes exemplars, transition matrices, the output probability distribution, and the prior probabilities. The exemplars are initialized as mentioned before and remain unchanged when estimating other parameters. The transition matrices and prior probabilities can be estimated using the expectation maximization (EM) algorithm. The exact forward–backward algorithm [27] is used in the E-step. The naive exact algorithm has the time complexity of $O(TK^{2M})$ if applying to the transformation of the factorial HMM into an equivalent HMM with K^m states using the forward–backward algorithm. 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 applied to get the most probable path and likelihood. New exemplars can be obtained through the most probable path and the new output probability distribution can be obtained from the new exemplars. The entire process is iterated until the likelihood converges to a small threshold.

3) Recognition: First, a probe cycle y is preprocessed and its features are extracted.

Then, the output probability distribution of the probe sequence can be calculated by the exemplars of the training sequence. We get the log likelihood P_j when the probe sequence is generated by the FHMM parameters λ_j of the jth person in the training database:

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

If P_m is the largest one among all P_j's, we assign the unknown person to be person m.

A key problem during calculating the log likelihood P_j is how to get the clusters of the probe sequence in case of the given parameters of the training sequence. Two methods can be considered. First, the distance between the features of the probe sequence and the exemplars of a training sequence are calculated to confirm the clusters. The clusters of the same probe sequence may vary with different training sequences. Second, the probe sequence is equally divided into K clusters, which remain the same for different train sequences. Our experiments adopt the second method.

C. PHMM-Based Silhouette Recognition

All the parameters are initialized and estimated by HMM instead of the FHMM. We use the frieze and wavelet features, respectively, to implement the HMM-based silhouette recognition classifiers and fuse their results. Both the voting strategies and score-based strategies [28] are adopted as the fusion rule. Two HMM classifiers vote for the final classification result. If they disagree with each other, the summation of the matching scores is realigned and used for the final classification. Therefore, the PHMM is considered as a kind of decision-level fusion.

IV. Numerical Experiments

We use the CMU MoBo gait database and CASIA gait database A to evaluate the proposed algorithm. The lateral sequences of the CMU MoBo gait database are adopted and the image size is preprocessed as 640*300. The CMU MoBo gait database contains 25 subjects. The lateral and oblique sequences of the CASIA gait database A are adopted and the image size is preprocessed as 120*95. The CASIA gait database A includes 20 subjects and four sequences for each viewing angle per subject.

The first part of the results is presented by CMS curves. In order to evaluate the algorithms comprehensively, we employ McNemar's test to obtain the second part of the results. For convenient description, some abbreviations are adopted hereafter. HMM(f) represents the HMM with frieze vectors. HMM(w) denotes the HMM with wavelet vectors. HMM(c) denotes the HMM with the concatenated vectors. The wavelet and frieze vectors of the same image are joined together to form a concatenated feature.

A. CMS Curves

The CMS curve indicates the probabilities of the correct match included in the top matches. The CMS curves of two gait databases are presented separately.

1) CMU MoBo Gait Database: Four experiments are performed. Four cycles are used for training and two cycles for testing. The experiments are set up as follows.

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

d) F versus B: Training on fast walk and testing on walking with a ball.

Fig. 6 gives the CMS curves of the four experiments using five different kinds of recognition methods. For the experiment "S versus F," as shown in Fig. 6(a), the FHMM method performs best whose gait recognition rate reaches 100% at rank 1. The results of HMM(f) and HMM(w) are good enough, but that of the PHMM method does not show much improvement. In experiment "F versus S" [see Fig. 6(b)], the feature concatenation method, i.e., HMM(c), gives the best result for all the rank values. The FHMM method performs close to the best one, but PHMM gives the lowest recognition rate at rank 1. For experiment "S versus B" [see Fig. 6(c)], the PHMM method has an advantage over others. However, this time the feature concatenation method HMM(c) has the lowest recognition rate for ranks 1–5. The HMM(w) and HMM(c) have comparative performances.

The results of experiment "F versus B" [see Fig. 6(d)] are more complex. For most rank values, the HMM(w) gives the highest recognition rate, while the HMM(f) gives the lowest one. This is because that when people walk with a ball, their shapes vary significantly. The frieze feature is more sensitive to the shape variation than the wavelet feature. The experimental results suggest that the frieze feature does not provide as much discriminate information as the wavelet feature to distinguish the gait state of fast walking and walking with a ball. Therefore, fusing the frieze feature with the wavelet feature either by FHMM, PHMM, or feature concatenation does not improve the recognition performance rather than by using the wavelet feature by itself. Another phenomenon in Fig. 6(d) is that the CMS curves of the PHMM and FHMM are intersect and it is difficult to tell which one performs better.

Generally speaking, FHMM is superior to HMM using a single feature under most circumstances. Also, the FHMM-based gait recognition method performs better than the PHMM-based method, except for experiment "S versus B." The PHMM also shows its advantages over single HMM, especially in experiments "S versus F" and "S versus B." It is worth noticing that when one feature performs badly, the performances of fusion are thereby degraded. HMM(c) performs poorly in the last two experiments, which indicates that it is more susceptible to shape changes.

2) CASIA Gait Database A: Four cycles are used for training and two cycles for testing. Lateral and oblique tests are carried out. The two tests are both trained on walking in one direction and tested on walking in the opposite direction.

Fig. 7 shows the CMS curves of the two experiments using five different kinds of recognition methods. However, the recognition rates are not noteworthy. The HMM(w) gives the best results while HMM(f) produces the worst. The results of FHMM and PHMM are similar, but worse than those of HMM(w). The fusion methods do not show any particular advantages.

We examine the images in the database and notice many incomplete silhouette images, which may affect the performance greatly. To deal with the incomplete gait image, we propose a new gait pattern representation method named FDEI. FDEI is represented as the sum of GEI and the difference image between

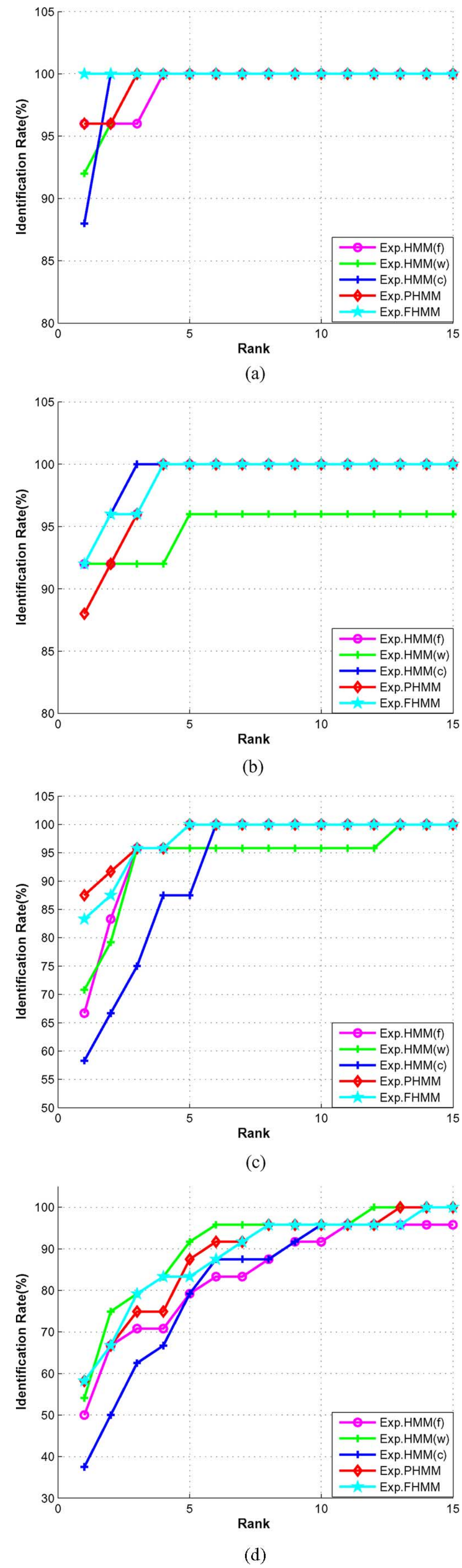
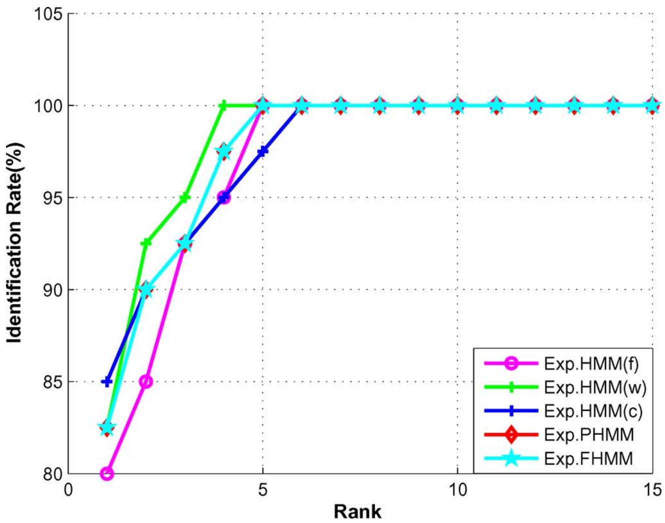
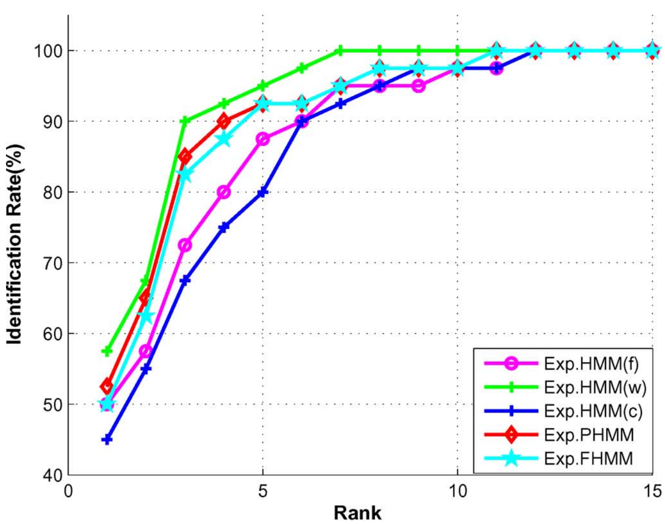


Fig. 6. CMS curves of the four experiments on the CMU MoBo gait database. (a) Results of S versus F. (b) Results of F versus S. (c) Results of S versus B. (d) Results of F versus B.



(a)



(b)

Fig. 7. CMS curves of two experiments on the CASIA gait database A. (a) Results of lateral test. (b) Results of oblique test.

the current frame and the next frame. The advantages of FDEI lie in emphasizing the difference between consecutive frames and maximally maintaining the primary shape information.

GEI is defined as follows [29]:

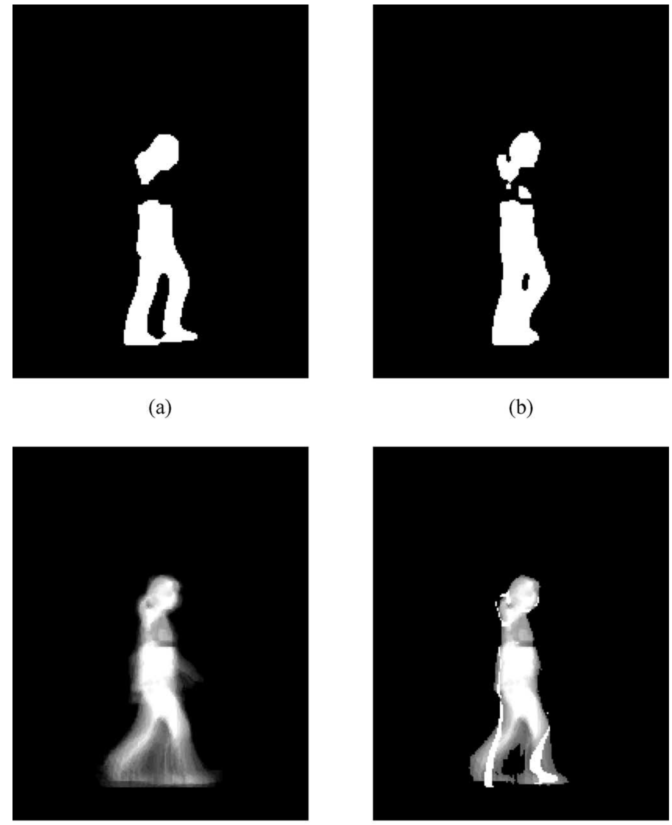
$$G(x,y) = \frac{1}{N}\sum_{t=1}^{N} B(x,y,t) \tag{7}$$

where N is the number of frames per cycle, t is the sequence number of the frame in the cycle (moment of time), and x and y are values in 2-D image coordinate. $B(x,y,t)$ is the silhouette image at time t. A GEI is shown in Fig. 8(c).

The FDEI of $B(x,y,t)$ can be represented as follows:

$$F(x,y,t) = B(x,y,t) - B\left(x,y,(t-1)\right) + G\left(x,y\right). \tag{8}$$

If $t = 1$, $B\left(x,y,(t-1)\right)$ is the last frame of the cycle. An example of the reprocessed image is displayed in Fig. 8. An FDEI is presented in Fig. 8(d).



(a) (b)

(c) (d)

Fig. 8. (a) Noisy silhouette at t. (b) Silhouette at $(t-1)$. (c) GEI. (e) FDEI of (a).

The CMS curves of the two experiments on the CASIA gait database A using the FDEI to extract the frieze feature are shown in Fig. 9. The performances of Fig. 9 are much better than that of Fig. 7. The frieze features extracted from FDEI improve the recognition rate significantly. For the lateral test [see Fig. 9(a)], FHMM gives the best performance. PHMM is the same with HMM(f) at the beginning, but it climbs slowly to rank 3. HMM(w) does not perform as well. HMM(c) is better than HMM(w). For the oblique test [see Fig. 9(b)], the performances of PHMM and FHMM are good. It is difficult to distinguish which one is better because they intersect. HMM(w) does not perform as well as HMM(f); however, it rapidly increases and is the first to achieve 100%. HMM(c) performs better than HMM(w), but worse than HMM(f). HMM(c) is more susceptible to the quality of the single feature than PHMM and FHMM.

From the two tests earlier on the CASIA gait database A, FDEI is verified to be an effective representation for the incomplete gait image. It should be noted that the advantages of the two fusion methods also rely on suitable feature selection.

B. McNemar's Test

The CMS curves given earlier show the cumulative results for ranks 1 and higher. However, the curves of the five methods always intersect and no single algorithm performs absolutely

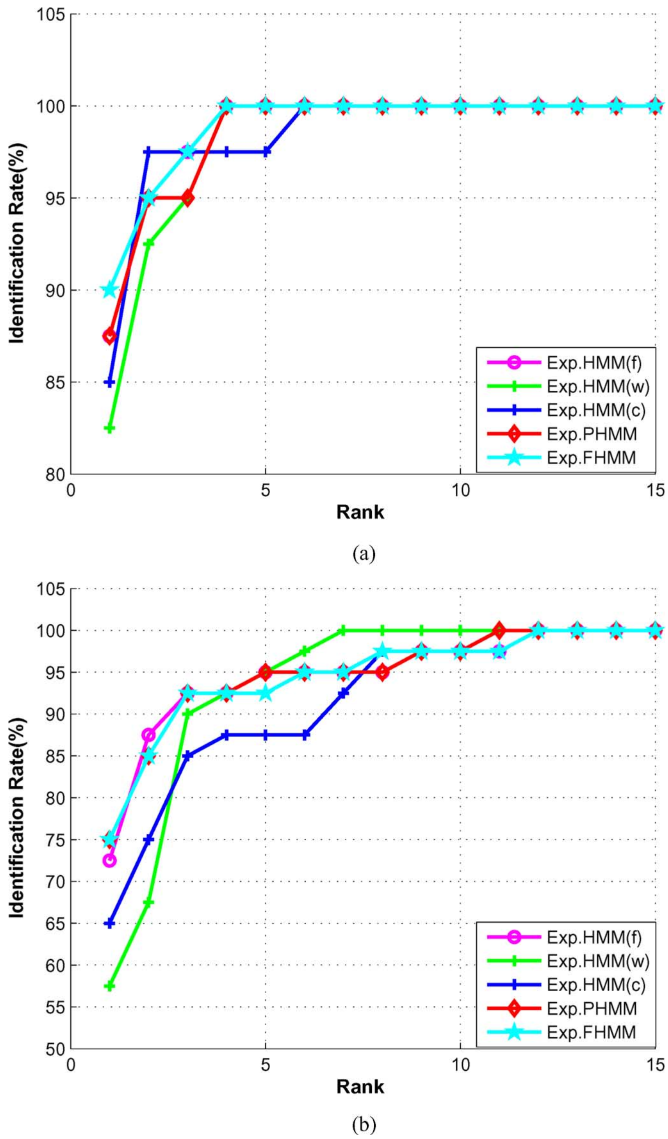


Fig. 9. CMS curves of two experiments on the CASIA gait database A using FDEI. (a) Results of lateral test. (b) Results of oblique test.

TABLE I
RESULTS OF MCNEMAR'S TEST

	Algorithm A Failed	*Algorithm A Succeeded*
Algorithm B Failed	N_{ff}	N_{sf}
Algorithm B Succeeded	N_{fs}	N_{ss}

better than others by comparing the CMS curves in different experiments. In order to comprehensively evaluate the algorithms better, we introduce the McNemar's test.

McNemar's test is a paired success/failure trial using the binomial model [30]. It is a first-order check on the statistical significance of an observed difference in recognition performance. Table I shows the results when two algorithms are compared.

TABLE II
PERFORMANCE COMPARISON OF HMM, PHMM, AND FHMM ON THE CMU MOBO GAIT DATABASE (ONE CYCLE TRAINING)

	N_{ss}	N_{sf}	N_{fs}	N_{ff}	Z-value	confidence
HMM(f) vs. HMM(w)	274	26	47	28	0.0274	51.20%
FHMM vs. HMM(f)	293	18	7	57	2.0000	97.72%
FHMM vs. HMM(w)	290	25	11	49	2.1667	98.46%
FHMM vs. HMM(c)	304	14	8	49	1.0660	85.54%
FHMM vs. PHMM	298	14	9	54	0.8341	79.67%

TABLE III
PERFORMANCE COMPARISON OF HMM, PHMM, AND FHMM ON THE CMU MOBO GAIT DATABASE (THREE CYCLES TRAINING)

	N_{ss}	N_{sf}	N_{fs}	N_{ff}	Z-value	confidence
HMM(f) vs. HMM(w)	2500	120	112	418	0.5249	69.85%
FHMM vs. HMM(f)	2597	137	42	374	6.6418	100%
FHMM vs. HMM(w)	2586	111	15	438	8.5714	100%
FHMM vs. HMM(c)	2841	35	28	246	0.8819	81.06%
FHMM vs. PHMM	2717	2	1	430	0	50%

McNemar's test is defined as

$$\chi^2 = \frac{(N_{sf} - N_{fs})^2}{(N_{sf} + N_{fs})}. \tag{9}$$

When $N_{sf} + N_{fs} < 40$, correction for continuity is needed. Therefore, the McNemar's test is corrected as

$$\chi^2 = \frac{(|N_{sf} - N_{fs}| - 1)^2}{(N_{sf} + N_{fs})}. \tag{10}$$

If the number of tests is greater than about 30, then the central limit theorem applies. In such a case, the Z score (standard score) is obtained from (10) as

$$Z = \frac{(|N_{sf} - N_{fs}| - 1)}{\sqrt{N_{sf} + N_{fs}}}. \tag{11}$$

If Algorithms A and B give very similar results, then Z will tend to zero. As their difference increases, Z will increase. Confidence limits can be obtained from the Z-value. More detail about McNemar's test can be found in [27].

1) CMU MoBo Gait Database: The previous experiments a) and b) of the CMU MoBo database are carried out. Tables II and III show the performance comparison results of HMM, PHMM, and FHMM. Table II gives the results of using one gait cycle for training and testing. Fifteen experiments are conducted for each person. Table III shows the results of using three cycles for training and testing. One hundred and twenty six experiments are done per person. After the numbers of N_{ss}, N_{sf}, N_{fs}, and N_{ff} are obtained, the Z-value is calculated as (11). We then look up the associate confidence limits in the standard normal school table.

Table II presents the results of using one fast/slow walk gait cycle for training and one slow/fast walk gait cycle for testing. It shows that FHMM is superior to HMM(f), HMM(w), HMM(c), and PHMM with confidence values of 97.72%, 98.46%, 78.67%, and 85.54%, respectively. The results indicate that fusing two features using factorial HMM not only produces better recognition performance than employing one feature only, but also produces better performances than the feature concatenation and parallel HMM fusion methods.

TABLE IV
PERFORMANCE COMPARISON OF HMM, PHMM, AND FHMM ON THE CASIA GAIT DATABASE A (TWO CYCLES TRAINING)

	N_{ss}	N_{sf}	N_{fs}	N_{ff}	Z-value	confidence
HMM(f) vs. HMM(w)	170	78	28	44	4.8564	100%
FHMM vs. HMM(f)	240	80	20	52	2.2678	98.81%
FHMM vs. HMM(w)	176	80	20	44	6.0000	100%
FHMM vs. HMM(c)	194	42	40	44	0.2208	58.71%
FHMM vs. PHMM	236	18	14	52	0.7071	75.80%

Another conclusion from Table II is that, according to the confidence value 51.20% of HMM(f) versus HMM(w), the frieze and wavelet features have similar discriminate abilities when one-cycle gait sequences are used for training and testing.

Table III shows the results of using three gait cycles for training and testing. One hundred and twenty six experiments are conducted per person. When more gait cycles are employed, the FHMM is absolutely better than the HMM(f) and HMM(w). FHMM is about 81.06% confidence superior to HMM(c). The feature concatenation method HMM(c) gives better performance than using the single feature, but is inferior to the PHMM and FHMM methods. Unlike the one-cycle experiments, FHMM ties with PHMM and HMM(f) showing about 69.85% confidence superior to HMM(w).

2) CASIA Gait Database A: The lateral and oblique tests are performed. Two cycles are used for training. Sixteen experiments are done per person. The results are illustrated in Table IV.

As revealed in Table IV, HMM(f) is 100% confidence superior to HMM(w), which mainly attributes to incomplete gait image. FHMM is about 98.81% confidence superior to HMM(f) and 100% confidence superior to HMM(w), which shows the advantage of fusing multiple features using factorial HMM. FHMM is 75.80% confidence superior to PHMM and 58.71% confidence superior to HMM(c). It can also be concluded that HMM(c) is somewhat better than PHMM and PHMM is better than HMM(f) and HMM(w). For space limitation, the results of PHMM versus HMM(c), HMM(f), and HMM(w) are not presented.

V. DISCUSSION AND CONCLUSION

The main contribution of this paper is that we employ the FHMM as feature-level fusion scheme to fuse different gait features. The PHMM decision-level fusion scheme is compared with FHMM. Experimental results show that the FHMM tends to perform better than PHMM when only a few gait cycles are available for recognition. FHMM tends to be easily implemented, while PHMM needs to obtain the results of the features separately and fuse them using certain rules. Besides PHMM, experiments with feature concatenation are included. The concatenated feature combined with the frieze and wavelet features is better than the single one, but inferior to the FHMM and PHMM. From the CMS curves, it can be concluded that the concatenated feature is more susceptible to varied shapes. Moreover, the concatenated feature may lead to the problem of dimensionality, the small sample size problem, and it may contain noisy or redundant data.

Another contribution of this paper lies in performance evaluation. The CMS curves usually show the results of the algorithms. However, the curves of different algorithms sometimes intersect and it is difficult to strictly decide which algorithm is better. Moreover, a single curve can only show the results of one experiment. In this study, we introduce McNemar's test to evaluate the methods statistically. It provides the statistical confidence limits for the comparison of two algorithms.

Besides these two contributions, FDEI is proposed to deal with incomplete human images. FDEI contains both the main shape of the person and the changes of the current frame. It complements the spatial information and preserves the temporal motion. The experimental results on the CASIA gait database A demonstrate FDEI is a good gait pattern representation.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the anonymous reviewers for their constructive comments. They would also like to thank Y. Liu and K. von Deneen for their help on improving the manuscript.

REFERENCES

[1] L. Wang, H. Ning, T. Tan, and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 14, no. 12, pp. 149–158, Feb. 2004.
[2] T. H. W. Lam, R. S. T. Lee, and D. Zhang, "Human gait recognition by the fusion of motion and static spatio-temporal templates," *Pattern Recognit.*, vol. 40, no. 9, pp. 2563–2573, 2007.
[3] A. I. Bazin, L. Middleton, and M. S. Nixon, "Probabilistic fusion of gait features for biometric verification," in *Proc. IEEE Int. Conf. Inf. Fusion*, Jul. 25–28, 2005, vol. 2, pp. 1211–1217.
[4] J. Han and B. Bhanu, "Statistical feature fusion for gait-based human recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*, Jun. 27–Jul 2, 2004, vol. 2, pp. 842–847.
[5] G. V. Veres, M. S. Nixon, L. Middleton, and J. N. Carter, "Fusion of dynamic and static features for gait recognition over time," in *Proc. IEEE Int. Conf. Inf. Fusion*, Jul. 25–28, 2005, vol. 2, pp. 1204–1210.
[6] A. Tyagi, J. Davis, and M. Keck, "Multiview fusion for canonical view generation based on homography constraints," in *Proc. ACM-MM Work. Video Surveillance Sens. Netw.*, Oct. 2006, pp. 61–69.
[7] A. Kale, A. Chowdhury, and R. Chellapa, "Towards a view invariant gait recognition algorithm," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS)*, Jul. 2003, pp. 143–150.
[8] Y. Wang, S. Yu, Y. Wang, and T. Tan, "Gait recognition based on fusion of multiview gait sequences," in *Proc. Int. Conf. Biometrics*, Jan. 2006, pp. 605–611.
[9] J. Lu and E. Zhang, "Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion," *Pattern Recognit. Lett.*, vol. 28, no. 16, pp. 2401–2411, 2007.
[10] Z. Liu and S. Sarkar, "Outdoor recognition at a distance by fusing gait and face," *Image Vis. Comput.*, vol. 25, no. 6, pp. 817–832, 2007.
[11] R. Chellappa, A. K. Roy-Chowdhury, and A. Kale, "Human identification using gait and face," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, Jun. 17–22, 2007, pp. 1–2.
[12] X. Zhou and B. Bhanu, "Feature fusion of face and gait for human recognition at a distance in video," in *Proc. IEEE Int. Conf. Pattern Recognit*, 2006, vol. 4, pp. 529–532.
[13] X. Zhou and B. Bhanu, "Integrating face and gait for human recognition at a distance in video," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1119–1137, Oct. 2007.
[14] X. Zhou and B. Bhanu, "Feature fusion of side face and gait for video-based human identification," *Pattern Recognit.*, vol. 41, no. 3, pp. 778–795, 2008.
[15] T. K. M. Lee, S. Ranganath, and S. Sanei, "Fusion of chaotic measure into a new hybrid face-gait system for human recognition," in *Proc. IEEE Int. Conf. Pattern Recognit*, Hong Kong, 2006, vol. 4, pp. 541–544.

[16] N. Boulgouris, D. Hatzinakos, and K. Plataniotis, "Gait recognition: A challenging signal processing technology for biometric identification," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 79–90, Nov. 2005.

[17] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H. Zhang, "Human gait recognition with matrix representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 896–903, Jul. 2006.

[18] B. Logan and P. J. Moreno, "Congestion avoidance in computer networks with a connectionless network layer," Cambridge Res. Lab., Cambridge, MA, Tech. Rep. CRL-97-7, Sep. 1997.

[19] Y. X. Liu, R. T. Collins, and Y. H. Tsin, "Gait sequence analysis using frieze patterns," in *Proc. Eur. Conf. Comput. Vis*, May 2002, pp. 659–671.

[20] R. Gross and J. Shi, "The CMU motion of body (MoBo) database," Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. RI-TR-01-18, Mar. 2001.

[21] CASIA Gait Database. (2004). [Online]. Available: http//www.cbsr.ia.ac.cn/english/Database.asp

[22] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," in *Proc. Conf. Adv. Neural Inf. Process. Syst. (NIPS)*, 1995, pp. 472–478.

[23] M. Brand, "Coupled hidden Markov models for modeling interacting processes," Massachusetts Inst. Technol. (MIT) Media Lab., Cambridge, Tech. Rep. 405, Jun. 1997.

[24] B. Logan and P. J. Moreno, "Factorial HMMs for acoustic modeling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process*, Seattle, WA, May 12–15, 1998, vol. 2, pp. 813–816.

[25] A. Betkowska, K. Shinoda, and S. Furui, "FHMM for robust speech recognition in home environment," in *Proc. Symp. Large-Scale Knowl. Resources*, Tokyo, Japan, 2006, pp. 129–132.

[26] A. Kale, N. Cuntoor, and R. Chellappa, "A framework for activity-specific human identification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process*, 2002, vol. 4, pp. 3660–3663.

[27] A. F. Clark and C. Clark. (2004). *Performance characterization in computer vision: A tutorial* [Online]. Available: http://peipa.essex.ac.uk/benchmark/

[28] B. Achermann and H. Bunke, "Combination of classifiers on the decision level for face recognition," Univ. Bern, Switzerland, Tech. Rep. IAM-96-002, Jan. 1996.

[29] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2002.

[30] J. R. Beveridge, K. She, B. A. Draper, and G. H. Givens, "Parametric and nonparametric methods for the statistical evaluation of human ID algorithms," Colorado State Univ., Fort Collins, Tech. Rep., Dec. 2001.

Changhong Chen received the B.S. degree in electronic engineering from Yantai University, Yantai, China, in 2004. She is currently working toward the Ph.D. degree in the School of Electronic Engineering, Xidian University, Xi'an, China.

Her current research interests include pattern recognition, image processing, and computer vision and their application in biometrics.

Jimin Liang received the B.S., M.S., and Ph.D. degrees from Xidian University, Xi'an, China, in 1992, 1995, 2000, all majored in electronic engineering.

In 1995, he joined Xidian University, where he is currently a Professor in the Life Science Research Center (LSRC), School of Electronic Engineering. During 2002, he was a Research Associate Professor in the Department of Electrical and Computer Engineering, University of Tennessee, Knoxville. His current research interests include information fusion, image processing, and pattern recognition.

Heng Zhao received the B.S. degree in automatic control from Xi'an Jiaotong University, Xi'an, China, in 1996, and the Ph.D. degree in circuit and system from Xidian University, Xi'an, in 2005.

From 1996 to 1999, he was an Assistant Researcher at the Flight Automatic Control Research Institute, Xi'an. He is currently an Associate Professor at the Life Science Research Center, School of Electronic Engineering, Xidian University. His current research interests include data mining, pattern recognition, image processing, and performance evaluation of video analysis algorithms.

Haihong Hu received the B.S. degree in image processing and transmission in 1994 from Beijing University of Post and Telecommunication, Beijing, China, and the M.S. degree in signal processing in 2001 from Xidian University, Xi'an, China, where she is currently working toward the Ph.D. degree in pattern recognition and intelligent system and works at the Life Science Research Center, School of Electronic Engineering.

Her current research interests include computer vision, machine learning, pattern recognition, and image classification, retrieval, and analysis.

Jie Tian (M'00–SM'03) received the Ph.D. degree (with honors) in artificial intelligence from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1992.

From 1994 to 1996, he was a Postdoctoral Fellow with the Medical Image Processing Group, University of Pennsylvania, Philadelphia. Since 1997, he has been a Professor at the Institute of Automation, Chinese Academy of Sciences. Since 2007, he has also been the Chair Professor of the Cheung Kong Scholars at Xidian University, Xi'an, China. His current research interests include pattern recognition, machine learning, image processing and their applications in biometrics, etc.