2-D Structure-Based Gait Recognition in Video Using incremental GMM-HMM

Rui Pu[1], Yunhong Wang[1]

[1]Laboratory of Intelligence Recognition and Image Processing, Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China

Abstract. Gait analysis is a feasible approach for human identification in intelligent video surveillance. However, the effectiveness of the dominant silhouette-based approaches are severely affected by dressing, bag, hair style and the like. In this paper, we propose a useful 2-D structural feature, named skeleton-based feature, effective improvements for human pose estimation in human walking environment and a recognition framework based on GMM-HMM using incremental learning, which can greatly improve the availability of gait traits in intelligent video surveillance. Our skeleton-based feature uses a 15-DOFs, which is effective in eliminating the interference of dressing, bag, hair style and the like, to represent the torso. In addition, to imitate the natural way of human walking, a Hidden Markov Model(HMM) representing the gait dynamics of human walking incrementally evolves from an average human walking model that represents the average motion process of human walking. Our work makes the gait recognition more robust to noise. Experiments on widely adopted databases prove that our proposed method achieves excellent performance.

1 Introduction

Gait, as a promising biometric characteristic, has attracted many researchers in recent years. In intelligent surveillance, the advantage of accessibility at a distance makes gait a promising biometric characteristic for human recognition. The silhouette has been regarded as the starting line of gait analysis because some databases provide silhouette directly and many gait researchers [1–3] managed to identify human by individual walking styles using silhouette-based methods. However, all the related methods are severely affected by dressing, bag, hair style and the like. Consequently, if someone changes his/her dressing or hair style, these methods perform badly. In this paper, we propose a new robust 2-D structural feature, effective improvements for human pose estimation in human walking environment and a recognition framework based on GMM-HMM using incremental learning. Furthermore, we assume that there is only one person walking in videos. Or if there are several persons superimposing each other, we cannot get skeleton-based feature, as a result, we cannot perform the identification.

In terms of feature, there have been some other efforts at gait analysis on 2-D structural feature. Guochang Huang [4] employed different blocks, which represent the solid silhouettes, and fitted the blocks with ellipse. Then, they perfomed recognition after merging the ellipse parameters of different view angles. Baofeng Guo [5] utilized the maximum mutual information (MMI) algorithm to select gait features, aiming at abandoning the redundant information in high-dimensional features and extracting the most important parts for identification. They applied the MMI to gait features, such as the size and position of each part of body and motion parameters like speed, then they performed recognition using Support Vector Machine (SVM). Their method achieved better performance than correlation analysis and variance analysis. But, in summary, these 2-D structural features are almost represented by shapes such as triangle, ellipse, polygon among others. Obviously, these shapes will be different when someone walks wearing thick clothes or carrying bags. Furthermore, these 2-D structural features are almost attained by background subtraction which is clumsy and rigorous to the video surveillance environment. On the other hand, skeleton-based feature is just human skeleton represented by 15-DOFs, which can reflect the eigen gait characteristics more thoroughly.

With regard to identification framework, there are several time series modeling methods such as Dynamic Time Wrap (DTW), Hidden Markov Model (HMM) and Conditional Random Field (CRF). First of all, DTW has a deadly limitation that it demands the same frequency between gallery set and probe set. Secondly, normal CRF is so complicated and unsuitable for gait analysis. Although the linear CRF is suitable, it is more sophisticated than HMM but not better than HMM on effectiveness for gait analysis. Consequently, we choose HMM. But, the normal HMM demands a mass of gait sequences as training samples. However, there are not enough samples in many cases. To conquer this problem, we get the individual HMMs evolving from an average HMM. In addition, the same person may walk at different time or different places under real circumstance so that the gait samples cannot be available in one shot, so the offline learning method is not suitable. On the contrary, incremental learning is rather useful to this problem. As a result, we get our incremental GMM-HMM evolving from an average GMM-HMM.

Incremental learning has been widely applied to many video-based applications, especially face tracking. For example, David A. Ross [6] proposed an online method based on incremental algorithm for Principal Component Analysis (P-CA). They updated the eigen dynamics using incremental learning. Most work mainly consider the variances of statistical features, since the motion dynamics seems less useful for recognition in their applications, such as face tracking. However, dynamics modeling is the core of gait analysis. In this paper, we attempt to incrementally learn the periodic gait dynamics, and exploit spatiotemporal relationships for recognition. Similar to [7], gait dynamics is regarded as the outward manifestation of stance transitions. Unlike some existing tracking methods such as particle filters [8] that depend on the similarities of appearances between frames, this work aims to recover and compare the periodic dynamics based on stance transitions. Furthermore, Maodi Hu [9] proposed an approach based on incremental learning, which achieved a good performance. An incremental learning method for HMM with Gaussian Mixture Model (GMM) representation (denoted as iGMM-HMM afterwards) is proposed, which shows promising performance in recognition experiments. The overall framework of the incremental learning process is shown in Figure 1.

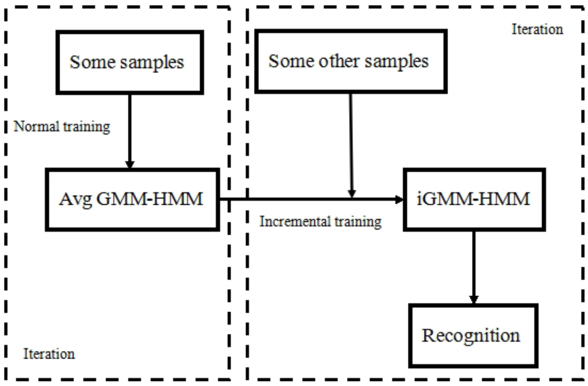
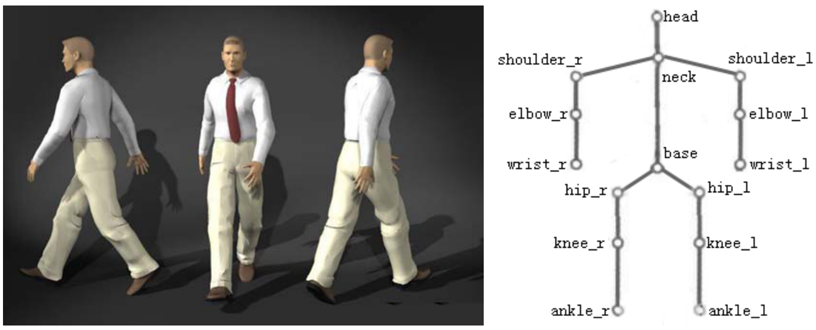


Fig. 1. Overall framework for incremental learning process

The remainder of this paper is organized as follows. Section 1.1 simply presents skeleton-based feature, human pose estimation method and its results. Section 2 is the technical details about human pose estimation method and iGMM-HMM. At last, section 3 is the experiment results in CASIA-B gait database.



(a) the normal human walking stance (b) the 15-DOFs we adopted

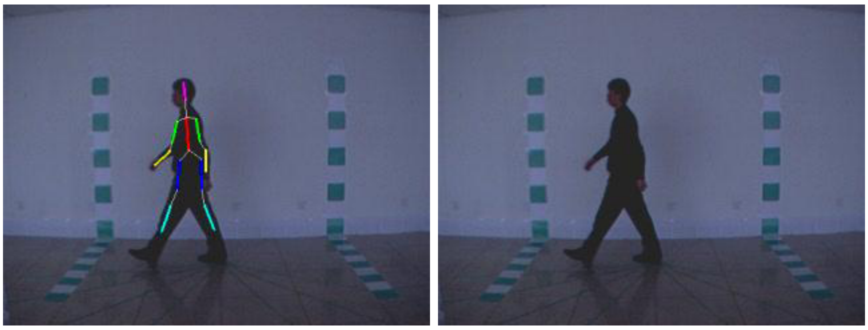
Fig. 2. the walking stance and the DOFs we adopted

1.1 Skeleton-based feature

As we know, human motion can be represented using Degree of Freedom nodes (DOFs) model. In this paper, we adopt the 15-DOFs, which is relatively easy to detect and track and enough to represent skeleton-based feature(see Fig. 2).

To get the 15-DOFs for each walking stance, there are two main approaches, which are human pose estimation and human pose tracking. Marcus A. Brubaker [10, 11] proposed a useful approach about human pose tracking for human in walking. But his method just tracks the lower body, which is insufficient to gait recognition. Furthermore, because the current human pose tracking methods are not good enough to achieve our goal and so complicated, we choose the first.

Vittorio Ferrari [12] proposed an approach for human pose estimation which achieved a good performance. But its method needs to label human upper body artificially. And then, he utilised another method [13] proposed by Navneet Dalal to detect upper body. But the upper-body detection method performed badly in low resolution images. Finally, Vittorio Ferrari [14] proposed a fully automated method for human pose estimation in uncontrolled environment. So we choose this method to be the base of our first part algorithm to get skeleton-based feature. In addition, their method is performed in still images, so we have to convert videos into images at first. Based on Ferrari's method, we made some improvements which can improve the upper-body detection accuracy and the image parsing speed. The difference between our's and Ferrari's is shown in(see Fig. 3).



(a) the human pose estimation result by our method

(b) because Ferrari's upper-body detection method cannot detect the upper body, nothing is attained

Fig. 3. the difference between our's and Ferrari's

The human pose estimation results by our method are shown in(see Fig. 4).

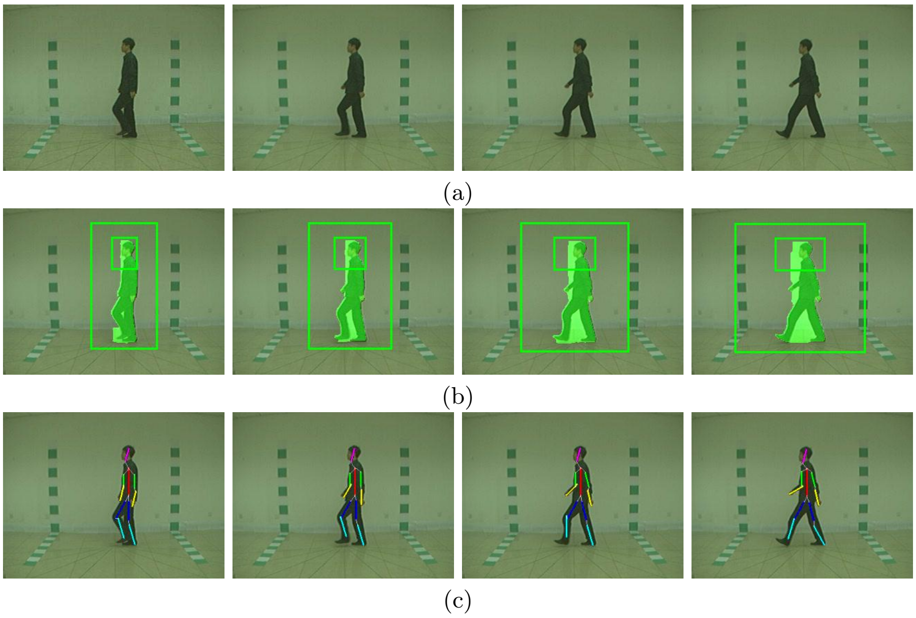


Fig. 4. The human pose estimation results step by step.(a) is the original images in gait database CASIA-B. (b) is the human detection results. (c) is the human pose estimation results.

2 Technical details

2.1 Skeleton-based feature

The main idea of the human pose estimation method proposed by Vittorio Ferrari is to progressively reduce the search space for body parts, greatly improving the chances that human pose estimation will succeed. In their approach, there are three stages in total.

1. Human detection. They started by detecting human upper bodies in every frame, using a sliding window detection based on Histograms of Oriented Gradients [13], and associate detections over time.

2. Foreground highlighting. At this stage, the search for body parts is limited in the detected regions.

3. Human pose estimation. They obtained a first pose estimation based on the image parsing technique of Ramanan [15]. The area to be parsed is restricted to the regions attained by foreground highlighting.

In their approach, the first two stages use a weak human model. This weak model only determines the approximate location and scale of the person, and roughly where the torso and head should lie. The last stage switches to a stronger model, a pictorial structure composed of body parts tied together in a tree-structured conditional random field. Parts, l_i are oriented patches of fixed size, and their positions are parameterized by location and orientation. The posterior of a configuration of parts $L = l_i$ given an image I can be written as a log-linear model

$$P(L|I) \propto exp(\sum_{(i,j)\in} E)\psi(l_i, l_j) + \sum_i \phi(l_i)) \qquad (1)$$

The binary potential $\psi(l_i, l_j)$ corresponds to a spatial prior on the relative positions of parts and embeds the kinematic constraints (e.g. the upper arms must be attached to the torso). The unary potential $\phi(l_i)$ corresponds to the local image evidence for a part in a particular position (likelihood). Since the model structure E is a tree, inference is performed exactly and efficiently by sum-product Belief Propagation.

Furthermore, there are some rules to utilise in the walking environment.

1. The upper arm must be above the lower arm, and the thigh must be above the shank.

2. The left arm and right leg, right arm and left leg must be in the same direction.

3. The left arm and right arm, left leg and right leg must be symmetric.

4. If the camera is fixed, the stance in the next frame must be near the pervious stance.

Based on their approach, by taking the advantage of the four rules, we make some effective improvements in the last stage.

First of all, according to the first three rules, we uses prior physiological characteristics of human gait and effectively limits the search space, which ameliorates the efficiency.

Secondly, in terms of the fixed camera, we initialize the position in current frame using the result of the previous one, and it only focuses on the person to be studied and largely reduces the number of false candidate, thus improving the performance.

Then, according to the human pose estimation results, we extracted skeleton-based feature, including the lengthes and angles of every two connective joints(see Fig. 2). The lengthes should be divided by the body height in the images. Since this feature is totally structural, it is hardly influenced by dressing, bag and hair style.

2.2 Incremental GMM-HMM

In gait analysis, stances are usually used to indicate the periodical latent states over gait cycles. After years of researches, human gait is widely accepted to be an identifiable periodic pattern with several stance phases. Consequently, a HMM that models the representation within and between states is very suitable for this application.

We will simply review the development of incremental learning for HMM below. Besides the offline Expectation Maximization (EM) algorithm and the batch learning Baum-Welch (BW) algorithm, the parameters of HMM can also be estimated incrementally with improved convergence and reduced memory requirements [16]. Krishnamurthy [17] derived online EM algorithm by using random approximations to maximize the Kullback-Leibler information. Stenger [18] proposed the Incremental Baum-Welch (IBW) algorithm, in which each latent state of their HMM includes a single Gaussian model. It is further derived to a discrete model with a new backward procedure based on a one-step lookahead by Florez-Larrahondo [16], which is known as the improved Incremental Baum-Welch (IBW+) algorithm, which achieved a better performance. In the purpose of learning gait dynamics for recognition, the models mentioned above should be enhanced. First, the model including the IBW+ is discrete. On the other hand, the model including the IBW involves only one Gaussian model for each latent state. But the model we need is continuous and may includes several Gaussian models for each state. Consequently, we apply the idea of IBW+ to our iGMM-HMM and learn the updating approach for GMM from the IBW.

About the symbol notation, we use iGMM-HMM to represent the incremental GMM-HMM we proposed and oGMM-HMM to represent the normal GMM-HMM gained by the offline BW algorithm. The feature vector extracted from t^{th} frame is indicated as O_t. There are some parameters in incremental learning. The Θ is used for the model representation, which is composed of the transition probability matrix A between latent states and the observable representations B. Each single stance within a gait cycle is represented by a latent state in HMM, and the probability density function (pdf) of each latent state is modeled by a GMM. Considering a HMM consisting of Q latent states with M Gaussian mixture components, $A = \{\alpha_{ij}\}_{1 \le i \le Q, 1 \le j \le Q}$ denotes the transition probability from latent state i to latent state j, and $B = \{\phi_{ik}, \mu_{ik}, \sigma_{ik}\}_{1 \le i \le Q, 1 \le k \le M}$ denotes the mixing coefficient, mean vector, and covariance matrix of component k in latent state i.

$\alpha_T(i) = P(O_1, \ldots, O_T, q_T = i|\Theta)$ is the forward cumulative probability of being in state i,

$$\alpha_T(i) = \begin{cases} (\sum_{j=1}^{Q} \alpha_{T-1}(j) a_{ji}) b_T(i) & T > 1, \\ b_T(i) & T = 1, \end{cases} \tag{2}$$

and $\beta_T(i) = P(O_T, O_{T+1}, q_T = i|\Theta)$ is the backward one proposed in IBW+ [16],

$$\beta_T(i) = \sum_{j=1}^{M} a_{ij} b_{T+1}(j). \tag{3}$$

Since the real $\beta_T(i)$ is based on an exponential decay function computed via the backward procedure, for large T this approximation seems to be appropriate. In any case, it provides a better approximation than $\forall_T \forall_i \beta_T(i) = 1.0$.

This backward procedure of IBW+ algorithm reduces the training complexity of β in backward procedure of BW algorithm in discrete model from $O(n^2T)$ to $O(n^2)$. Although it does not improve the global time complexity, the experimental results in [16] show that IBW+ converges faster than BW and IBW. Note that it requires a one-step look ahead in the sequence of observations.

$b_T(i) = P(q_T = i|O_T, \Theta)$ is the pdf of O_T at state i, which indicates the fitness of a single frame for an averaged walking stance. Because of the usage of

IBW+, both $b_T(i)$ and $b_{T+1}(i)$ are updated in the T^{th} iteration.

$$b_T(i) = \sum_{k=1}^{M} \phi_{ik} \mathcal{N}(O_T; \mu_{ik}, \sigma_{ik}), \tag{4}$$

$$b_{T+1}(i) = \sum_{k=1}^{M} \phi_{ik} \mathcal{N}(O_{T+1}; \mu_{ik}, \sigma_{ik}). \tag{5}$$

$c_T(i,k)$ is the probability of O_T being in component k at state i,

$$c_T(i,k) = \frac{\phi_{ik} \mathcal{N}(O_T; \mu_{ik}, \sigma_{ik})}{b_T(i)}, \tag{6}$$

$\gamma_T(i) = P(q_T = i | O_1, \ldots, O_{T+1}, \Theta)$ is the probability of being in state i,

$$\gamma_T(i) = \frac{\alpha_T(i)\beta_T(i)}{\sum_{i=1}^{Q} \alpha_T(i)\beta_T(i)}, \tag{7}$$

$\xi_{T-1}(i,j) = P(q_{T-1} = i, q_T = j | O_1, \ldots, O_{T+1}, \Theta)$ is the probability of $T-1^{th}$ frame being in state i and Tth frame being in state j,

$$\xi_{T-1}(i,j) = \begin{cases} \frac{\alpha_{T-1}(i) a_{ij} b_T(j) \beta_T(j)}{\sum_{i=1}^{Q} \sum_{j=1}^{M} \alpha_{T-1}(i) a_{ij} b_T(j) \beta_T(j)} & T > 1, \\ 0 & T = 1. \end{cases} \tag{8}$$

The estimation of $\xi_{T-1}(i,j)$ is improved by the approximation of β_T [16]. We use the parameters of an average GMM-HMM (denoted as avgGMM-HMM afterwards) to serve as the model representation Θ of the iGMM-HMM in the 0^{th} iteration. Given the T^{th} and $T+1^{th}$ frames, $b_T(i)$, $b_{T+1}(i)$, $c_T(i,k)$, $\alpha_T(i)$, $\beta_T(i)$, $\gamma_T(i)$, and $\xi_{T-1}(i,j)$ can be calculated in order, based on Θ in the $T-1^{th}$ iteration.

Below we will introduce our incremental updating algorithm. At first, supposing there are N frames in the training group of avgGMM-HMM, we number them as $x_{-N+1}, \ldots, x_0$ to differentiate them from the frames in incremental learning. Given the values of model parameters estimated in the previous frames, the equations suitable for T^{th} updating are shown in Equation (9) to (12).

$$\overline{a}_{ij}^{T} = \frac{\overline{a}_{ij}^{T-1}(\sum_{t=-N+1}^{T-2}\gamma_t(i)) + \xi_{T-1}(i,j)}{\sum_{t=-N+1}^{T-1}\gamma_t(i)}, \tag{9}$$

$$\overline{\phi}_{ik}^{T} = \frac{\sum_{t=-N+1}^{T}\gamma_t(i)c_t(i,k)}{\sum_{t=-N+1}^{T}\gamma_t(i)}, \tag{10}$$

$$\begin{aligned}\overline{\mu}_{ik}^{T} =& \frac{\overline{\mu}_{ik}^{T-1}(\sum_{t=-N+1}^{T-1}\gamma_t(i)c_t(i,k))}{\sum_{t=-N+1}^{T-1}\gamma_t(i)c_t(i,k) + \gamma_T(i)c_T(i,k)} \\ &+ \frac{\gamma_T(i)c_T(i,k)O_t}{\sum_{t=-N+1}^{T-1}\gamma_t(i)c_t(i,k) + \gamma_T(i)c_T(i,k)},\end{aligned} \tag{11}$$

$$\begin{aligned}\overline{\sigma}_{ik}^{T} =& (\overline{\sigma}_{ik}^{T-1} + (\overline{\mu}_{ik}^{T-1} - \overline{\mu}_{ik}^{T})(\overline{\mu}_{ik}^{T-1} - \overline{\mu}_{ik}^{T})^H) \\ &\cdot \frac{\sum_{t=-N+1}^{T-1}\gamma_t(i)c_t(i,k)}{\sum_{t=-N+1}^{T-1}\gamma_t(i)c_t(i,k) + \gamma_T(i)c_T(i,k)} \\ &+ \frac{\gamma_T(i)c_T(i,k)(O_T - \overline{\mu}_{ik}^{T})(Ox_T - \overline{\mu}_{ik}^{T})^H}{\sum_{t=-N+1}^{T-1}\gamma_t(i)c_t(i,k) + \gamma_T(i)c_T(i,k)},\end{aligned} \tag{12}$$

Compared to previous studies on incremental HMM [16, 18], such as IBW and IBW+, the proposed updating rules make it possible to model the state representations of the HMM by several Gaussian models.

3 Our experiment

Recognition approaches based on HMM is straight-forward. Let Θ^{id} denote the HMM trained by the gallery set of subject *id*. Given the data-case *probe*, the recognition process can be simply solved by Maximal A Posterior (MAP) rule,

$$\text{argmax}_{id} P(probe|\Theta^{id}). \tag{13}$$

where $P(probe|\Theta^{id})$ is the probability of the observation sequence *probe* given Θ^{id}.

3.1 The database introduction

The database we used is CASIA-B gait database. There are 124 persons in total, 11 view angles for each person, three types for each view angles. The view angles are 0, 18, 36, 54, 72, 90, 108, 126, 144, 162 and 180 degree respectively. The types are nm, bg and cl, respectively standing for dressing normally, wearing thick clothes and carrying bag. There are only two gait sequences for bg and cl and six for nm.

Before we choose to use the iGMM-HMM, we attempted to build oGMM-HMM for each person, each angle and each type. In a small probe set including 622 gait sequences, we experimented and concluded that the oGMM-HMM for each type performs better than that for each angle and each type. The experimental results are shown in Table 1. The person oGMM-HMM, angle oGMM-HMM and type oGMM-HMM stand for the oGMM-HMM for each person, angle and type respectively.

Table 1. Rank 1 recognition performance with view angle unknown(%).

Approaches	person oGMM-HMM	angle oGMM-HMM	type oGMM-HMM
Accuracy	74.65%	76.85%	80.06%

Consequently, building GMM-HMM for each type is the best. However, there are only two gait sequences for bg and cl and six for nm in CISIA-B gait database. And also we have to extract at least one gait sequence for each type as probe set, so that the training samples are too small to make the GMM-HMM convergent. As a result, the initial parameter settings are far away from the true values, then the errors will slow down the convergence process [18]. Therefore, we trained an average GMM-HMM using some gait samples, whose parameters are estimated using the offline EM algorithm and the BW algorithm. Then, with incremental adjustments of the iGMM-HMM parameters, the fitness and validity of specific individuals increase simultaneously.

3.2 The contrastive methods we used

In our experiment, except for our proposed method, we also take three other methods as contrastive methods.

Skeleton-based feature plus sub-sequence DTW. The traditional dynamic time warping (DTW) algorithm is to compute the distance from the probe sequence to the gallery sequence. But in many cases, we need the minimum distance from the sub-sequences of the probe sequence to the sub-sequences of the gallery sequence, so that we shouldn't compute the distance from the probe sequence to the gallery sequence. In our experiment, for simplicity, we make sure that the probe sequence is shorter than the gallery sequence. Consequently, we just need to compute the minimum distance from the probe sequence to the sub-sequences of the gallery sequence. So, we call this method sub-sequence DTW(denoted as subDTW afterwards). The recognition results are based on the distance between the *gallery* set and the *probe* one.

$$\mathrm{argmin}_{id} d(probe, gallery_{id}) \quad (14)$$

$d(S_1, S_2)$ represents the Euclidean distance between two sequences denoted by S_1 (in the gallery set) and S_2 (in the probe set). Let T_1 and T_2 denote the lengths of S_1 and S_2 respectively.

$$d(S_1, S_2) = \min_{s=1}^{T_1-T_2+1} || \sum_{t_1=s}^{s+T_2} S_1(t_1) - \sum_{t_2=1}^{T_2} S_2(t_2) ||. \tag{15}$$

Skeleton-based feature plus offline GMM-HMM. The traditional oGMM-HMM for each type using Skeleton-based feature may not converge because of the small amount of samples. In this method, we initialize the prior probability and the initial transition probability from latent states to observable states with uniform distribution. In addition, the initial transition possibility from latent states to latent states is stochastic.

Gait energy image (GEI) plus PCA plus nearest-neighbor classifier. This method uses the classical feature GEI. Then, after PCA process, the nearest-neighbour classifier can achieve a very good performance. This method is denoted as GEI-PCA-NN afterwards.

3.3 Experimental results

In our experiment, we choose 40 persons randomly, including 20 women and 20 men respectively, to train the avgGMM-HMM(only the iGMM-HMM uses this). Then we split the left gait sequences into probe set and gallery set. The probe set includes one gait sequence for each type and the left is the gallery set. The results are shown in Table 2.

In terms of the parameters setting, there are two controllable parameters which are the number Q of latent states and the number M of Gaussian mixture components for each latent state. The experiments prove that iGMM-HMM performs better when Q is between 6 to 8 and M is between 3 to 4.

Obviously, the iGMM-HMM is better than the oGMM-HMM and subDTW with the same skeleton-based feature we proposed. Furthermore, although iGMM-HMM plus skeleton-based feature performs a little worse than GEI-PCA-NN in nm-nm, bg-bg and cl-cl, the results with cumulative match score(CMS)(see Fig. 5) prove that iGMM-HMM catches up with GEI-PCA-NN quickly. Anyway, the recognition accuracy of iGMM-HMM is still high in nm-nm, bg-bg, cl-cl. In addition, iGMM-HMM performs a little bit worse in bg-cl, which is mainly because the human pose estimation results are relatively worse in bg and cl than nm. But in the majority of cross-type recognitions such as nm-bg, nm-cl, bg-nm and so on, iGMM-HMM is obviously better than GEI-PCA-NN, moreover, the superiority keeps the same with CMS. In a word, our skeleton-based feature is a better feature than GEI in cross-type recognition. However, the cross-type recognition accuracy is still a little low, especially in bg-cl, cl-nm and cl-bg, whose reason is that there are still some errors in the human pose estimation method.

Table 2. Rank 1 recognition performance with view angle unknown in type to type(%)

Approaches	subDTW	oGMM-HMM	iGMM-HMM	GEI-PCA-NN
nm-nm	95.89	93.71	98.53	**98.83**
nm-bg	47.80	49.27	**70.97**	53.67
nm-cl	30.79	28.74	**43.99**	26.10
bg-nm	39.31	38.71	**53.96**	36.66
bg-bg	88.27	84.59	91.79	**95.60**
bg-cl	16.13	17.72	17.89	**20.23**
cl-nm	22.58	20.31	**23.17**	17.60
cl-bg	16.67	16.42	**20.23**	12.90
cl-cl	83.87	81.33	86.80	**97.07**

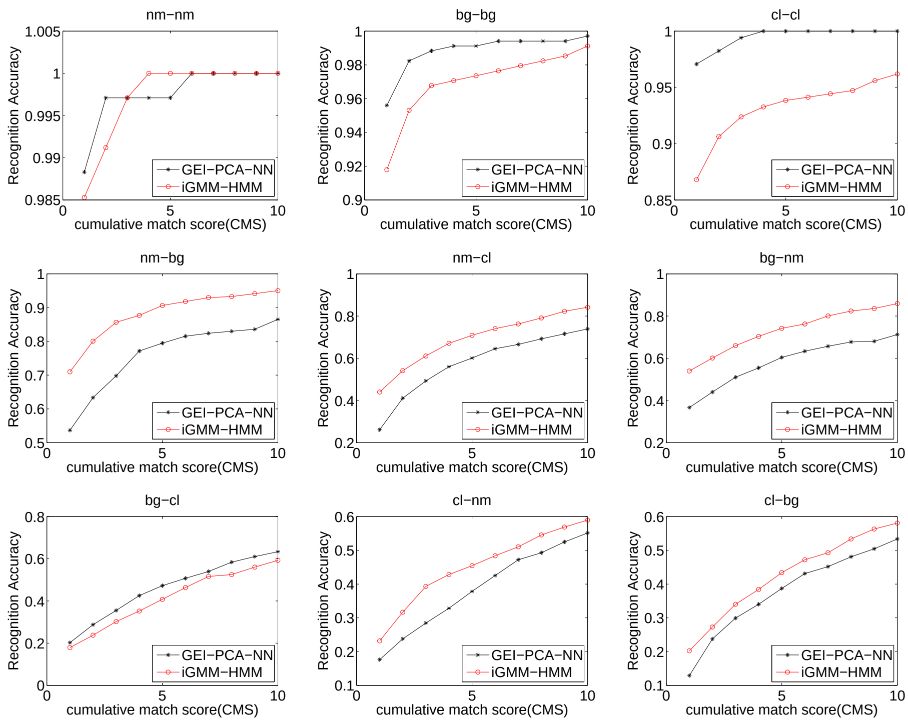


Fig. 5. Comparison between iGMM-HMM and GEI-PCA-NN in CMS

4 Conclusions and future work

In this paper, a novel 2-D structural feature, effective improvements for human pose estimation in human walking environment and an incremental identification framework for gait dynamics are proposed. The experiments prove that our skeleton-based feature can eliminate the interference of dressing, bag, hair style and the like effectively. However, only structural feature is not enough to human identification problem in gait analysis. As a result, our future work should be fusing the skeleton-based feature with some other features to cover this shortage. In addition, in spite of the improvement we do in human pose estimation, the human pose estimation results are still not that much good because of some detection errors. But, as the human pose estimation or human pose tracking improves, our approach must achieve better performance. Whatsoever, the iGMM-HMM is really a good framework for spatiotemporal problems.

References

1. Liu, Z., Malave, L., Sarkar, S.: Studies on silhouette quality and gait recognition. In: CVPR. (2004)
2. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: Averaged silhouette. In: ICPR. (2004)
3. Yu, S., Tan, T., Huang, K., Jia, K., Wu, X.: A study on gait-based gender classification. IEEE J IP **18** (2009) 1905–1910
4. Huang, G., Wang, Y.: Gender classification based on fusion of multi-view gait sequences. In: ACCV. (2007) 462–471
5. Guo, B., Nixon, M.S.: Gait feature subset selection by mutual information. IEEE transactions on systems, man, and cybernetics **39** (2009) 36–46
6. Ross, D.A., Lim, J., Lin, R.S.: Incremental learning for robust visual tracking. IJCV **77** (2008) 125–141
7. Liu, Z., Sarkar, S.: Improved gait recognition by gait dynamics normalization. IEEE J PAMI **28** (2006) 863–876
8. Ristic, B., Arulampalam, S., Gordon, N.: Beyond the kalman filter: Particle filters for tracking applications. In: Artech House. (2004)
9. Hu, M., Wang, Y., Zhang, Z., Zhang, D., Little, J.J.: Incremental learning for video-based gait recognition with lbp flow. IEEE TRANSACTIONS ON CYBERNETICS **43** (2013)
10. Brubaker, M.A., Fleet, D.J., Hertzmann, A.: Physics-based person tracking using simplified lower-body dynamics. In: CVPR. (2007)
11. Brubaker, M.A., Fleet, D.J.: The kneed walker for human pose tracking. In: CVPR. (2008)
12. Ferrari, V., Zisserman, A.: Progressive search space reduction for human pose estimation. In: CVPR. (2008)
13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR. (2005)
14. Ferrari, V., Zisserman, A.: 2d human pose estimation in tv shows. In: Dagstuhl. (2009)
15. DevaRamanan: Learning to parse images of articulated bodies. In: NIPS. (2006)

16. Florez-Larrahondo, G., Bridges, S., Hansen, E.A.: Incremental estimation of discrete hidden Markov models based on a new backward procedure. In: AAAI. Volume 1. (2005) 758–763
17. Krishnamurthy, V., Moore, J.B.: On-line estimation of hidden markov model parameters based on the kullback-leibler information measure. IEEE J SP **41** (1993) 2557–2573
18. Stenger, B., Ramesh, V., Paragios, N., F.Coetzee., J.Buhmann: Topology free hidden markov models: Application to background modeling. In: ICCV. Volume 1. (2001) 294–301