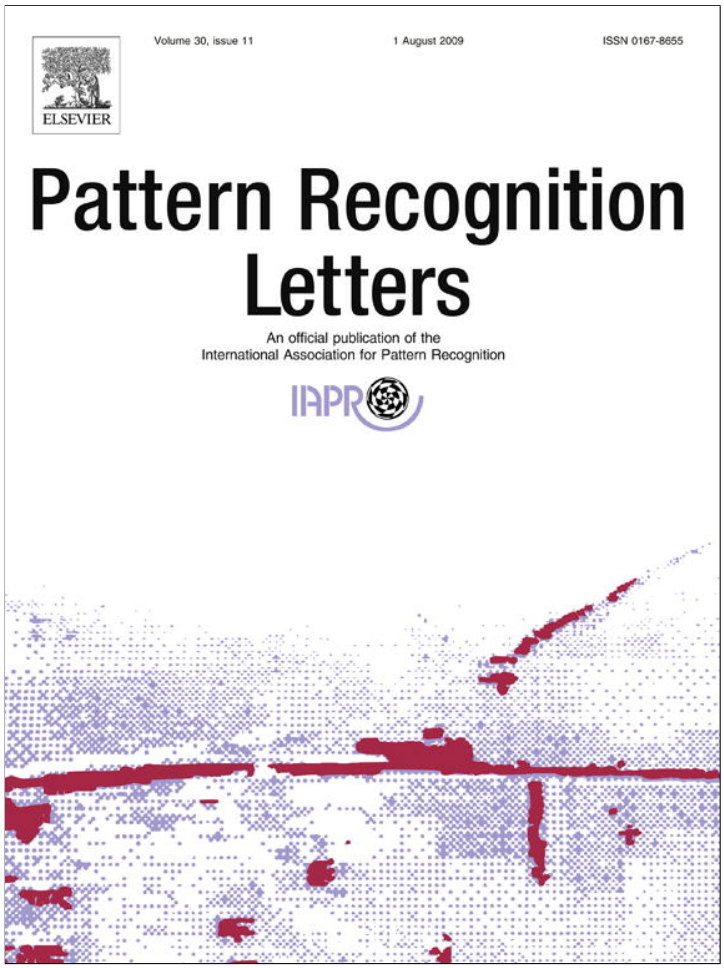

Pattern Recognition Letters 30 (2009) 977–984

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Frame difference energy image for gait recognition with incomplete silhouettes

Changhong Chen [a], Jimin Liang [a,*], Heng Zhao [a], Haihong Hu [a], Jie Tian [a,b]

[a] Life Science Research Center, School of Electronic Engineering, Xidian University, Xi'an, Shaanxi 710071, China
[b] 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 100190, China

ARTICLE INFO

Article history:
Received 19 August 2008
Received in revised form 22 February 2009
Available online 5 May 2009

Communicated by N. Pears

Keywords:
Gait recognition
Incomplete silhouettes
Frame difference energy image
Hidden Markov model

ABSTRACT

The quality of human silhouettes has a direct effect on gait recognition performance. This paper proposes a robust dynamic gait representation scheme, frame difference energy image (FDEI), to suppress the influence of silhouette incompleteness. A gait cycle is first divided into clusters. The average image of each cluster is denoised and becomes the dominant energy image (DEI). FDEI representation of a frame is constructed by adding the corresponding cluster's DEI and the positive portion of the frame difference between the former frame and the current frame. FDEI representation can preserve the kinetic and static information of each frame, even when the silhouettes are incomplete. This proposed representation scheme is tested on the CMU Mobo gait database with synthesized occlusions and the CASIA gait database (dataset B). The frieze and wavelet features are adopted and hidden Markov model (HMM) is employed for recognition. Experimental results show the superiority of FDEI representation over binary silhouettes and some other algorithms when occlusion or body portion lost appears in the gait sequences.

1. Introduction

Gait recognition aims to identify people at a distance by the way they walk. It usually comprises four steps: video capture, silhouette segmentation, feature extraction and recognition. Given a video of a subject walking across the image plane, we are only interested in the foreground walking subject. In order to eliminate the impact of background and color or texture of the clothes, the walking subject needs to be segmented from the background and represented as a binary silhouette. The common segmentation strategies rely on background subtraction (Elgammal et al., 2000; Friedman and Russell, 1997) or grouping optic flow to find the coherent motion (Tian and Shah, 1996; Yacob and Black, 1999). A human body could not always be perfectly segmented from the background, even if the video were of relatively good quality. Many factors lead to inaccuracy in the silhouette segmentation, such as the similarity of colors of the subject and background, illumination change, moving objects in the background, the distance between the camera and the person, occlusions and so on. These factors lead to spurious pixels, shadows, holes inside moving subject, noisy contours and incomplete silhouettes (body portion lost). These low quality silhouettes need further processing to achieve high recognition performance. When the silhouettes have small noisy areas or holes, they can be removed by some morphological operations, such as dilation and erosion. Some algorithms deal with more complicated noise, which can be divided into three categories.

The first category is the reconstruction-based method. Liu et al. proposed population hidden Markov models (HMM) to reconstruct the silhouette (Liu et al., 2004; Liu and Sarkar, 2005). The population HMM helps to map a frame in any given sequence to a stance and an appearance-based Eigen-Stance model is used to reconstruct the computed silhouette in the frame. This method can reconstruct silhouettes visually appealing and robust to viewpoint variation, but the lost of characteristics possessed by a single subject deteriorate recognition performance.

The second category focuses on aligning the contours. Yu et al. made use of improved dynamic time warping (IDTW) to reduce the effect of noise on human contours (Yu et al., 2007). A point on a contour is aligned to multiple points on another contour using traditional DTW. Only the pair with the shortest distance is kept and other pairs which have a common point are discarded in the IDTW. This method improves the recognition performance.

The third category is the robust static representation, which compresses a gait cycle into one or several static images. Han and Bhanu proposed the gait energy image (GEI), which is the average image of a gait cycle to characterize human walking properties (Han and Bhanu, 2006). When the noise at different moments is uncorrelated and identically distributed, GEI was found to be less sensitive to silhouette noise in individual frames. The performance of GEI is notable, but this representation loses detailed information and does not contain temporal variation. The gait history image (GHI) (Liu and Zheng, 2007) and gait moment image (GMI)

* Corresponding author. Tel./fax: +86 29 88202649.
E-mail address: jimleung@mail.xidian.edu.cn (J. Liang).

0167-8655/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2009.04.012

(Ma et al., 2007) were developed based on GEI. GHI preserves the temporal information besides the spatial information. It overcomes the shortcoming of no temporal variation in GEI. However, each cycle only obtains one GEI or GHI template, which easily leads to the problem of insufficient training cycles. GMI is the gait probability image at each key moment of all gait cycles. The corresponding gait images at a key moment are averaged as the GEI of this key moment. However, it is not easy for GMI to select key moments from cycles with different periods.

The low quality silhouette segmentation with body portion lost deteriorates the recognition performance. The reconstruction-based methods use common characteristics of the population to reconstruct visual appealing silhouettes, however, it does not improve the recognition performance. The method of aligning the contours fails when considerable shape change occurs. The static representations are robust to shape change in some frames, but they lose too much dynamic and detailed information. In this paper, a robust dynamic gait representation scheme, the frame difference energy image (FDEI), is proposed to suppress the effect of silhouette incompleteness. A gait cycle is divided into clusters and the dominant energy image (DEI) is obtained by denoising the averaged image of each cluster. The frame difference is calculated by subtracting two consecutive frames. The FDEI representation of a frame is constructed as the summation of its corresponding cluster's DEI and the positive portion of its frame difference, which embodies both static and kinetic information. To evaluate the new representation, experiments are carried out on the CMU Mobo gait database (Gross and Shi, 2001) with synthesized occlusions and the CASIA gait database (dataset B). Frieze feature (Liu et al., 2002) and wavelet feature are extracted from the original binary silhouette and the FDEI representation, respectively. HMM is used to train models and gives the recognition results.

The rest of this paper is organized as follows. In Section 2, we describe the silhouette preprocessing and the proposed FDEI gait representation in detail. Section 3 introduces the HMM-based gait recognition method. The experiment results on the CMU Mobo gait database with synthesized occlusions and the CASIA gait database (dataset B) are presented and analyzed in Section 4. Section 5 offers our conclusion.

2. Frame difference energy image (FDEI) representation

The FDEI representation is constructed on the preprocessed silhouette sequences. In this section, we first introduce the silhouette extraction and preprocessing methods and then present how to construct the FDEI representation.

2.1. Silhouette extraction and preprocessing

The silhouette is usually extracted by finding the difference between the background and current image (Elgammal et al., 2000; Friedman and Russell, 1997) or grouping optic flow to find the coherent motion (Tian and Shah, 1996; Yacob and Black, 1999). There are inevitably spurious pixels, holes inside moving subject and other anomalies in the detected sections. Mathematical morphological operations, such as erosion and dilation, are widely used to remove spurious pixels and fill small holes inside the extracted silhouettes. To eliminate the size difference caused by the varying distance between the subject and camera, the silhouettes are usually height scaled and centered. Even if the original images were of good visual perception, some inaccuracy may not be repaired by mathematical morphological operations. Silhouette incompleteness caused by body portion lost has a greater effect on recognition than other errors, such as shadows and spurious pixels. When the heads or feet are missing, height scaling will cause serious shape distortion. Fig. 1 shows some sample images of the background images, original images, raw silhouettes and preprocessed silhouettes. The first row of Fig. 1 is sample images of CMU Mobo gait database and the other three rows show sample images of CASIA gait database (dataset B). The first and second rows are good silhouette extraction sample images. The last two rows illustrate imperfect silhouette extraction and the preprocessed silhouette in the fourth row appears with a large shape distortion after aligning the height.

2.2. FDEI representation construction

Gait recognition depends on two types of information: shape information (static information) and dynamic information. Shape information refers to the appearance, such as body height, width, body-part proportions, hair, hunching, dressing, taking a package or ball and so on. Some shape information remains unchangeable or similar during people walking. Generally speaking, information extracted from a single frame belongs to shape information. Dynamic information characterizes the motion changes of the human body regardless of the underlying structure, such as joint-angles trajectories of the lower limbs (Wang et al., 2004), shape-of-motion features obtained by optical flow (Little and Boyd, 1998) and so on. Experiments with portions of the average silhouette representation showed that recognition power is not entirely derived from upper body shape; rather the dynamics of the legs also contribute equally to recognition (Nixon and Carter, 2006).

An effective gait representation should embody shape and dynamic information and be robust to imperfect silhouettes. Silhouette incompleteness tends to be more intractable and harmful than other errors. Here, we propose a new robust dynamic representation scheme, the FDEI. The main motivation of the proposed FDEI representation is to suppress the effect of the silhouette incompleteness while keeping most of the shape details and con-

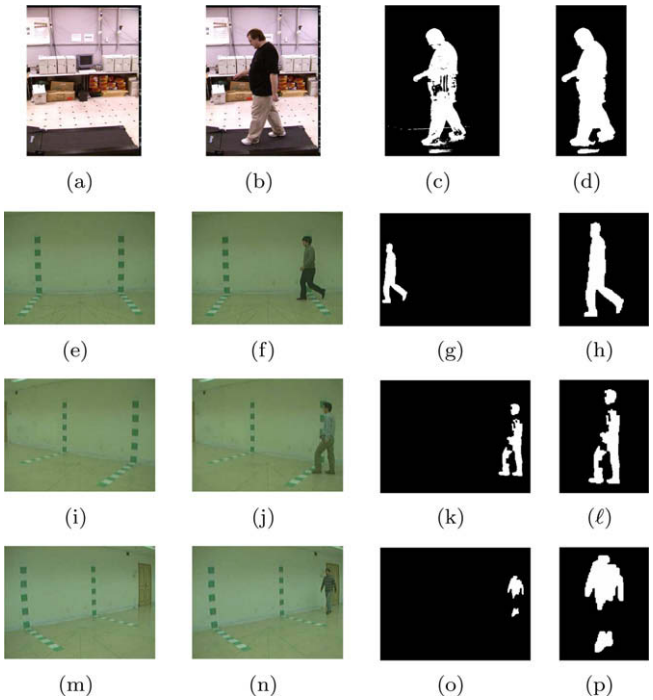

Fig. 1. Sample images of human silhouettes and their preprocessed silhouettes: the first column is of the background images; the second column is of the original images; the third column is the raw silhouettes; the fourth column is of the preprocessed silhouettes.

taining the temporal variation. The incompleteness of a certain frame cannot be repaired by itself but may be compensated by other frames to recover the shape information. We first divide a gait cycle into clusters and denoise the averaged image of each cluster to become the DEI. In order to characterize each frame, the frame difference is calculated as the subtraction of two consecutive frames. The positive portion of the frame difference contains the motive pixels and any incomplete parts. The FDEI representation of a frame is constructed as the summation of its corresponding cluster's DEI and the positive portions of its frame difference.

We follow the steps below to construct the FDEI representation of a gait sequence.

Step 1: Clustering and calculating GEI. The silhouettes of a gait cycle are clustered. The cluster number is chosen according to the average distortion (Kale et al., 2002), which decreases with the increased cluster number. If the average distortion does not change appreciably beyond certain number, it can be selected as the cluster number. When representing the gait cycle with smaller number clusters, some useful information loses. The larger cluster number leads to great computation complexity, but has little improvement on the recognition performance. Any suitable clustering algorithm can be used. We simply divide the silhouettes into temporally adjacent clusters of approximately equal number of frames. Different clusters have different stance, such as feet adjacent, toe off, heel rise and so on. The GEI of the cth cluster, $G_c(x,y)$, is computed as Han and Bhanu (2006):

$$G_c(x,y) = \frac{1}{N_c}\sum_{t\in A_c} B(x,y,t), \tag{1}$$

where t is moment of time, A_c represents the time set of silhouettes in the cth cluster, N_c is the number of frames in the cth cluster, x and y are coordinates in the 2D image and $B(x,y,t)$ is the silhouette at time t.

Step 2: Denoising. The GEI of the cth cluster $G_c(x,y)$ is denoised as:

$$D_c(x,y) = \begin{cases} G_c(x,y), & \text{if } G_c(x,y) \geqslant T, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $D_c(x,y)$ is the denoised image of $G_c(x,y)$. The threshold T varies with different cycles or subjects, depending on the quality of the silhouettes. The quality of the silhouettes is difficult to be measured automatically. Also it is not practical to manually label it for each cluster of each gait cycle. We experimentally choose the value T as $max(G_c) * 0.8$ for each cluster in the following gait recognition experiments. We name the denoised image $D_c(x,y)$ as the DEI.

Step 3: Calculating the positive portion of frame difference. The frame difference is computed as the subtraction of silhouettes $B(x,y,t-1)$ and $B(x,y,t)$. The positive portion $F(x,y,t)$ is obtained by setting the negative pixel values of the frame difference to zero, as expressed in Eq. (3)

$$F(x,y,t) = \begin{cases} 0, & \text{if } B(x,y,t) \geqslant B(x,y,t-1), \\ B(x,y,t-1) - B(x,y,t), & \text{otherwise}. \end{cases} \tag{3}$$

When $t = 1$, $B(x,y,t-1)$ is set to the last frame of the cycle. If $B(x,y,t)$ is incomplete and $B(x,y,t-1)$ is complete, $F(x,y,t)$ will contain the missing and movement portions. Otherwise, it only embodies the movement portions.

Step 4: Constructing FDEI. We define the FDEI as the summation of the positive portion of the frame difference and the corresponding cluster's DEI, as expressed in Eq. (4). The summation $FD(x,y,t)$ is regarded as the FDEI representation of $B(x,y,t)$

$$FD(x,y,t) = F(x,y,t) + D_c(x,y). \tag{4}$$

When $B(x,y,t)$ is incomplete and $B(x,y,t-1)$ is complete, the incomplete portions of the frame are contained in $F(x,y,t)$. When both $B(x,y,t)$ and $B(x,y,t-1)$ are incomplete, $D_c(x,y)$ can partially compensate the missing portions. The FDEI representation helps to suppress the effect of the missing portions and preserve the characteristics of $B(x,y,t)$. Fig. 2 demonstrates some images during the construction of FDEI, when both $B(x,y,t)$ and $B(x,y,t-1)$ are incomplete. The first two images of Fig. 2 show silhouettes $B(x,y,t)$ and $B(x,y,t-1)$, respectively. The positive portion of the frame difference (Fig. 2c) shows the movement portion of $B(x,y,t)$. The GEI of the cluster containing $B(x,y,t)$ is shown as Fig. 2d, which contains some weak information. The DEI (Fig. 2e), which embodies the dominant energy of the GEI, is obtained by denoising the GEI. Fig. 2f is the FDEI representation of $B(x,y,t)$. Comparing Fig. 2f with Fig. 2a, it can be seen that the FDEI representation contains the movement portion and partially compensates the incompleteness of $B(x,y,t)$.

3. HMM-based gait recognition

HMM has been the dominant technology in speech recognition since the 1980s'. HMM provides a very useful paradigm to model the dynamics of speech signals. Gait is similar with speech in time-sequential space. HMM was introduced to gait recognition in recent years (Kale et al., 2002; Sundaresan et al., 2003) and gained inspiring performance. Since then, more gait recognition papers (Liu et al., 2004; Chen et al., 2006; Cheng et al., 2008) based on HMM have appeared. 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 (Boulgouris et al., 2005).

Fig. 3 shows the topological representation of an HMM. The parameters of HMM $\lambda = (A, B, \pi)$ are initialized and trained as mentioned in (Chen et al., 2006). The whole recognition process contains three steps:

Step 1: Initializing. The feature vectors $\{f_t\}$ of each cycle are divided into C clusters. e_c denotes the center of the cth cluster C_c and is regarded as an exemplar. The initial transition probability matrix $A = \{a_{i,j}, i = 1\cdots C, j = 1\cdots C\}$ only allows current state equally transforming to itself and its next state, that is, $a_{i,i} = a_{i,i+1} = 0.5$. The observation symbol probability $B = \{b_c(f_t)\}$ is an exponential

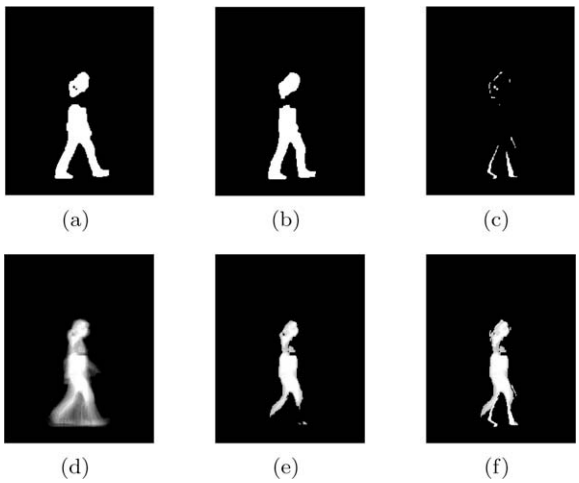

Fig. 2. Sample images during the construction of FDEI: (a) an incomplete silhouette at t; (b) the silhouette at $t-1$; (c) the positive portion of the frame difference; (d) the GEI; (e) the DEI; (f) the FDEI of (a).

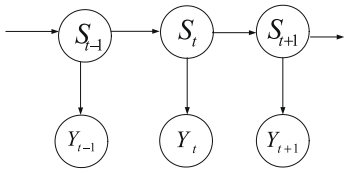

Fig. 3. Topological representation of an HMM.

function of the inner product distances of the feature vectors and the exemplars

$$b_c(f_t) = \alpha\delta_c e^{-\delta_c D(f_t, e_c)}, \tag{5}$$

where $D(f_t, e_c)$ is the inner product distance of f_t and e_c, α is a constant less than 1 and δ_c is defined as:

$$\delta_c = \frac{N_c}{\sum_{f_t \in C_c} D(f_t, e_c)}, \tag{6}$$

where N_c is the number of frames of the cth cluster C_c. The initial probabilities π_c are set to be $\frac{1}{N_c}$.

Step 2: Training. Viterbi algorithm is used to get the most probable path, from which the new exemplars are obtained. Then the new observation symbol probability is calculated from Eqs. 5 and 6. Given the new observation symbol probability, the new transition probability and initial probability can be updated using Baum-Welch algorithm. This process iterates a few times to obtain the acceptable parameters.

Step 3: Recognition. The observation symbol probability of the probe sequence is calculated using the exemplars of the training sequence. Other parameters retain the same with the training sequence. The similarities are calculated using Viterbi algorithm. The recognition result corresponds to the largest similarity.

4. Numerical experiments

The FDEI representation aims to reduce the influence of incomplete binary silhouettes. In order to evaluate the new representation method, we tested it on the indoor CMU Mobo gait database with synthesized occlusion and the CASIA gait database (dataset B). Two features are separately extracted to train the HMM. The first one is the frieze feature, which is constructed by stacking row projections (Liu et al., 2002). The frieze feature $F(y,t)$ of binary silhouette $B(x,y,t)$ is defined as:

$$F(y,t) = \sum_x B(x,y,t). \tag{7}$$

The frieze feature is smoothed to eliminate spurious pixels. Fig. 4 shows the frieze feature of Fig. 2f and its smoothed one. Another adopted feature is the wavelet feature. A two-dimensional wavelet transform is applied to the images using Haar wavelet bases. The wavelet coefficients of the approximation sub-image include the most useful information and are chosen as the wavelet feature.

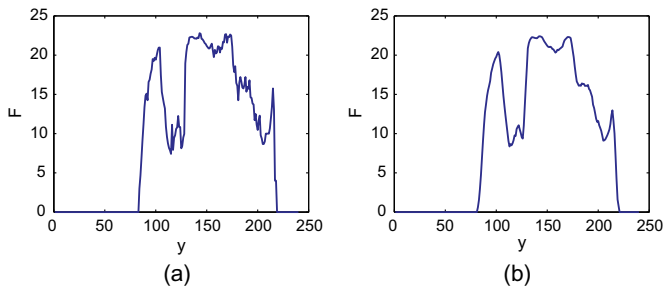

Fig. 4. The curves of the frieze feature: (a) the frieze feature of Fig. 2f; (b) smoothed frieze feature of (a).

4.1. CMU Mobo gait database with synthesized occlusion

The CMU Mobo gait database consists of sequences from 25 subjects walking on a treadmill, positioned in the middle of a room. Each subject is recorded performing four different types of walking: slow walk, fast walk, slow walk holding a ball, and a walk on an inclined plane. Each sequence is recorded 11 s long, recorded at 30 frames per second. The sequences caught by the frontal-view camera are adopted in following experiments. Fast walk sequences are chosen as the gallery and slow walk sequences as the probe set.

We simulate the occlusion situations by adding horizontal or vertical bars to the gallery. The bars have the same color as the background. A horizontal or vertical bar is added to a gallery silhouette with the probability varying from 10% to 100%. The adding bar probability represents the percentage of frames with bars in the gait sequence and is denoted as BP hereafter in the figures. The position of the added bar is uniformly distributed within the silhouette height or width. Fig. 5 displays sample images of simulated horizontal and vertical occlusions with variable bar width.

After adding bars to the gallery, gait recognition experiments based on HMM are conducted to compare the FDEI representation and the original binary silhouette. Two features, the frieze and wavelet features, are extracted from the FDEI and the original silhouette, respectively. Therefore, four experiments are carried out, i.e., frieze feature from binary silhouette (frieze(BI)), frieze feature from FDEI (frieze(FD)), wavelet feature from binary silhouette (wavelet(BI)) and wavelet feature from FDEI (wavelet(FD)). Given the adding bar probability and bar width, we repeat each experiment 5 times and the recognition results are averaged to reduce the statistical error.

Fig. 6 shows the identification rates when adding horizontal bars. The horizontal bar width varies from 40 to 100 pixels with step size of 20 pixels. The recognition rates of adding vertical bars are shown in Fig. 7, where the vertical bar width changes from 20 to 50 pixels with step size of 10 pixels. For a given bar width, the adding bar probability BP increases from 10% to 100%. The average recognition rates over BP are presented in Tables 1 and 2. Based on Figs. 6, 7, Tables 1 and 2, we come to the following conclusions.

First of all, for the synthesized occlusion data, gait recognition using the FDEI representation directly outperforms the original

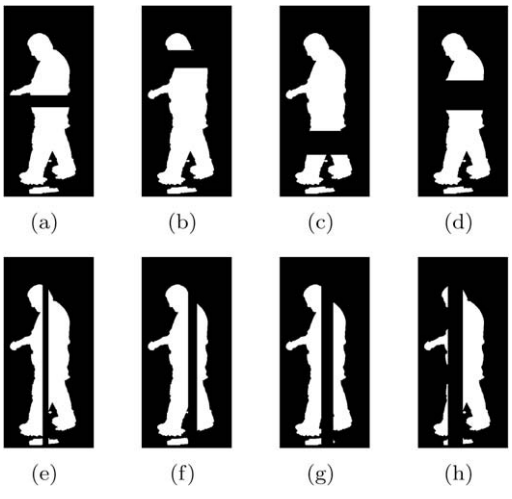

Fig. 5. Illustration of synthesized occlusions: (a)–(d) Horizontal bars with width of 40, 60, 80 and 100 pixels; (e)–(h) Vertical bars with width of 20, 30, 40 and 50 pixels.

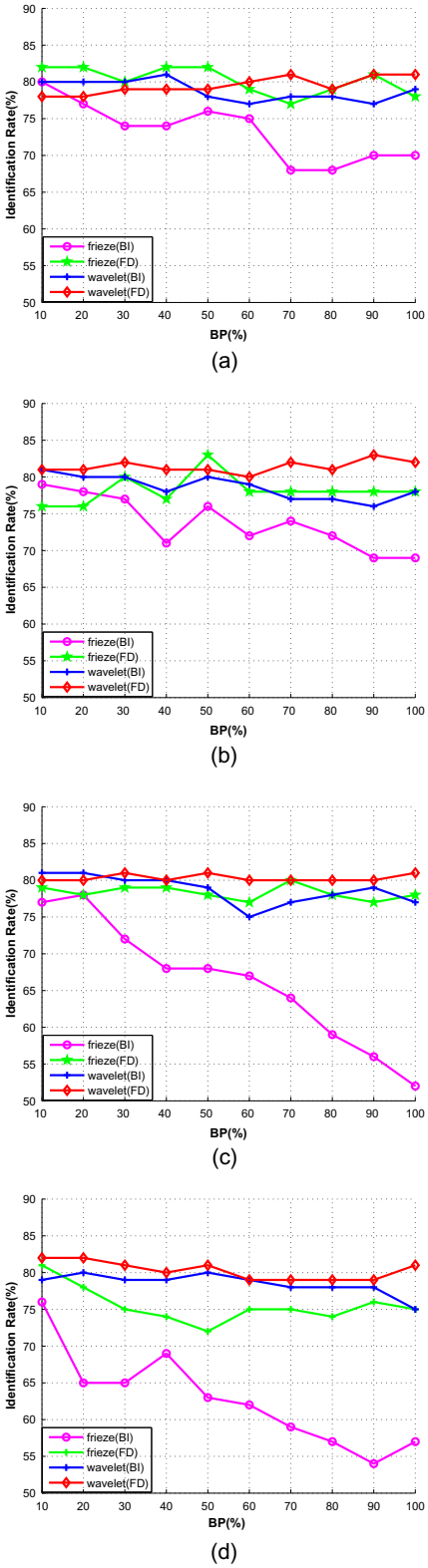

Fig. 6. Identification rate vs. adding horizontal bars probability: (a)–(d) represent the curves with a bar width of 40, 60, 80 and 100 pixels, respectively.

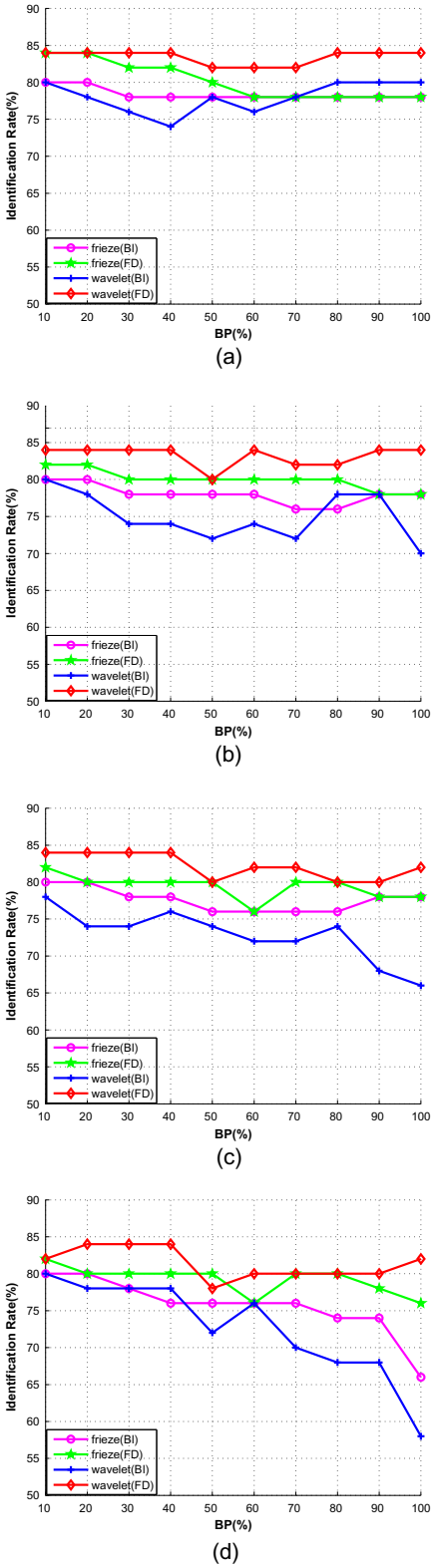

Fig. 7. Identification rate vs. adding vertical bars probability: (a)–(d) represent the curves with a bar width of 20, 30, 40 and 50 pixels, respectively.

binary silhouette. This is clearly demonstrated in Tables 1 and 2 where all the average recognition rates of FDEI are higher than their counterparts of using the original binary silhouette, despite the feature or the occlusion bar width. Fig. 6 shows that, under the horizontal bar occlusion situations, the FDEI representation with frieze feature or wavelet feature (frieze(FD) or wavelet(FD)) has similar performance to the original binary silhouette using wavelet feature (wavelet(BI)). However, recognition using the binary silhouette with frieze feature is greatly deteriorated by the occlusions. For the vertical occlusions, as shown in Fig. 7, the FDEI

Table 1
The average recognition rate (%) vs. horizontal bar width.

	40		60		80		100	
	BI	**FD**	BI	**FD**	BI	**FD**	BI	**FD**
Frieze	73.2	**80.2**	73.7	**78.2**	66.1	**78.3**	62.7	**75.5**
Wavelet	78.8	**79.5**	78.6	**81.4**	71.6	**80.3**	78.5	**80.3**

Table 2
The average recognition rate (%) vs. vertical bar width.

	20		30		40		50	
	BI	**FD**	BI	**FD**	BI	**FD**	BI	**FD**
Frieze	78.4	**80.2**	78.0	**80.0**	77.6	**79.4**	75.6	**79.2**
Wavelet	78	**83.4**	75	**83.2**	72.8	**82.2**	72.6	**81.4**

representation is obviously superior than the binary silhouette, regardless of the frieze or wavelet features used.

Secondly, the wavelet feature performs better than the frieze feature under most circumstances, especially when using the FDEI representation. The wavelet feature describes the holistic information of the original information and tends to be more credible. However, its recognition performance is more sensitive to the adding probability of vertical bars than that of the frieze feature.

Thirdly, the recognition rates fluctuate and do not decrease absolutely when the adding bar probability increases. However, the wider the occlusion, the FDEI representation shows more superiority to the binary silhouettes. When the width of the horizontal bars increases, the average recognition rates of frieze(BI) drop from 73.2% to 62.7% and those of frieze(FD) fluctuate around 78%. When vertical bars are added, the average recognition rates of wavelet(BI) vary from 72.6% to 78% and those of wavelet(FD) vary from 81.4% to 83.4%.

We also compare the proposed method with GEI, GHI, GMI and IDTW, whose results are shown in Tables 3 and 4. Linear discriminant analysis is used to calculate the recognition results of GEI, GHI and GMI. GEI performs much better than GHI and GMI for both horizontal and vertical bars. Although IDTW is proved to reduce the effect of noise on human contour, this algorithm does not perform well with synthesized occlusion. When horizontal or vertical bars are added, the contour is composed of two separate parts and IDTW cannot efficiently reduce the effect of noise under such circumstance. The performance of wavelet(FD) is the best among the five algorithms.

4.2. CASIA gait database

The proposed FDEI representation is also tested on CASIA gait database (dataset B), which contains 124 subjects (93 males and 31 females) captured from 11 view angles. There are six normal walking sequences for each subject per view. Many silhouettes in CASIA gait database (dataset B) are incomplete after background subtraction. We visually examine the database to get a quantitative evaluation of the silhouette quality. To alleviate labor intensity, only the gait sequences of angle 36 are inspected. We visually check the silhouettes after cycle detection and count the number of severely incomplete silhouettes. We only count the silhouettes lost head or feet portion because, following the aligning operation, these incomplete silhouettes cause considerable shape distortions. The silhouettes lost body portion but kept their height are not taken into account. Some incomplete examples are shown in the first row of Fig. 8. Their corresponding foregoing frames and FDEIs are illustrated in the second and third rows of Fig. 8, respectively.

We visually inspect the gait silhouette quality of the CASIA database. For simplicity, only view angle 36 is evaluated and the result is shown in Table 5. The number of severely incomplete frames are counted and divided by the training frames to find the percentage of incomplete frames. Among the 124 people, only 30 people have less than 5% incomplete frames. About half of the people have 10–25% incomplete frames. Only 14 people have more than 30%

Table 3
The average recognition rate (%) of different algorithms vs. horizontal bar width.

Method	40	60	80	100
IDTW	64	60.2	62.4	63.2
GEI	79.6	80.6	81	79.6
GHI	54.4	54.4	57.8	53.4
GMI	46.0	46.4	46.4	39.6
Wavelet(FD)	**79.5**	**81.4**	**80.3**	**80.3**

Table 4
The average recognition rate (%) of different algorithms vs. vertical bar width.

Method	20	30	40	50
IDTW	66.2	67.3	65.8	66.4
GEI	81	82	82	80.6
GHI	52.8	54.6	56	56.2
GMI	48.8	50.8	46.4	48.4
Wavelet(FD)	**83.4**	**83.2**	**82.2**	**81.4**

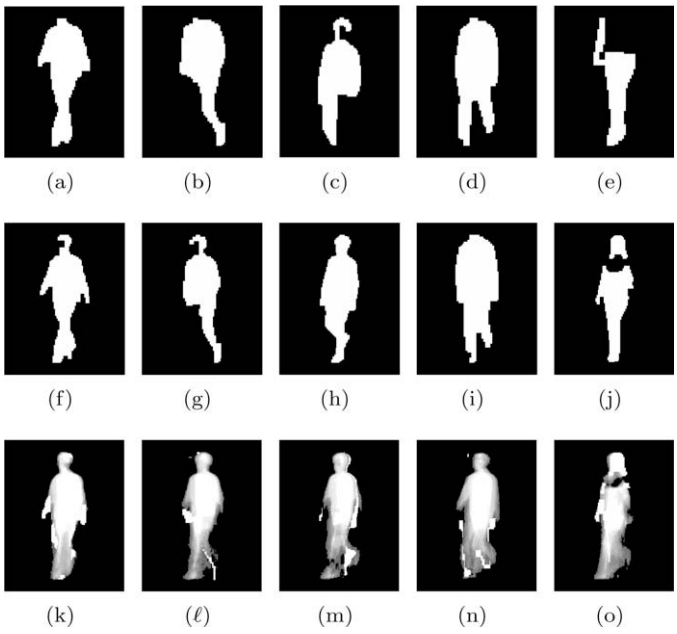

Fig. 8. Examples of incomplete silhouettes in CASIA gait database and their FDEI: (a)–(e) examples of incomplete silhouettes; (f)–(j) the foregoing frames corresponding to the silhouettes in the first row; (k)–(o) the FDEI of the silhouettes in the first row.

Table 5
The silhouette quality evaluation for view angle 36 of the CASIA gait database (dataset B).

Percentage of incomplete frames	Subject number
0–5	30
5–10	31
10–15	23
15–20	16
20–25	10
25–80	14

Table 6
The average recognition rates (%) of all the view angles.

Method	0	18	36	54	72	90	108	126	144	162	180	**Avg.**
IDTW	93.5	84.7	85.9	80.2	83.9	83.5	73.0	80.6	89.9	90.7	92.7	**85.3**
GEI	91.1	86.3	71.8	66.1	90.3	83.1	78.2	79.0	90.3	88.7	89.9	**83.2**
GHI	71.8	62.1	48.4	41.9	80.4	71.8	54.0	64.5	66.1	54.0	68.5	**62.1**
GMI	68.5	56.4	52.4	48.4	66.1	60.4	54.0	55.6	51.6	59.6	62.9	**57.8**
Frieze(BI)	95.2	61.3	48.4	48.4	58.1	79.8	64.5	49.2	55.6	71.8	84.7	**65.2**
Frieze(FD)	**95.2**	**84.7**	**84.7**	**93.5**	**90.3**	**91.1**	**86.3**	**84.7**	**85.5**	**88.7**	**95.2**	**89.1**
Wavelet(BI)	89.5	83.1	75.8	73.4	78.2	77.4	77.4	70.2	70.2	75.8	85.5	**77.9**
Wavelet(FD)	**100**	**100**	**100**	**93.4**	**81.1**	**90.3**	**90.3**	**86.3**	**91.9**	**91.9**	**97.6**	**93.9**

incomplete frames. When the incomplete frames happen inconsecutively in the gait sequence, they do not affect recognition much. However, in practice, incomplete frames often appear consecutively, which may severely deteriorates recognition performance.

The experiments are carried out for all view angles of the CASIA gait database (dataset B). The first four sequences are used for training, and the last two are placed into the probe set. The results are also compared with GEI, GHI, GMI and IDTW. Table 6 shows the correct classification rates (CCRs) at rank 1. The performance at angles 0 and 180 is better than at other angles. GEI performs much better than GHI and GMI, but little worse than IDTW. Wavelet(FD) has the best recognition rates for most angles and reaches 100% for the first three angles. Frieze(FD) performs better than IDTW except at angles 36, 144 and 162. IDTW performs better than wavelet(BI) except at angle 108 and frieze(BI) shows the worst performance. The cumulative match score (CMS) curves are illustrated in Fig. 9, which give the average CCR of all angles from rank 1 to rank 10. Most of the curves are not crossed except those of IDTW and GEI. It can be seen that none of these algorithms performs better than FDEI, no matter frieze or wavelet feature are extracted.

In order to obtain the quantitative superiority of the proposed FDEI representation over binary silhouette, we further employ McNemar's test. 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 limits and produce the evaluation results. McNemar's test is generally more comprehensive and reliable than the CMS curve to compare two algorithms. The two given algorithms, A and B, are compared. Four numbers, N_{ss}, N_{sf}, N_{fs} and N_{ff}, can be obtained, which represent the times of both algorithms succeed, algorithm A succeeds but B fails, algorithm A fails but B succeeds and both algorithms fail, respectively. Then a Z-value is calculated as follows:

$$Z = \frac{(|N_{sf} - N_{fs}| - 1)}{\sqrt{N_{sf} + N_{fs}}}. \tag{8}$$

Confidence limits can be obtained by looking up the Z-value in the standard normal distribution table. Detailed information of McNemar's test refers to Clark and Clark (1999).

Gait sequences of three view angles, 72, 90 and 162, are chosen and 15 experiments are carried out for each of the three angles per person. The wavelet and frieze features are used separately. The performances of McNemar's test are displayed in Table 7. It can be obtained from the table that the FDEI representation performs absolutely better than the binary silhouettes, despite the angle or feature. Although the confidence limits are all 100%, their difference is embodied in the Z-values. A higher Z-value of the FDEI representation indicates its superiority over the binary silhouette under all circumstances.

The foregoing numerical experiments testify that the FDEI representation is robust to the incompleteness of the gait silhouettes. Compared with some other algorithms, it has much better performance. The results also illustrate that the wavelet feature performs better than the frieze feature.

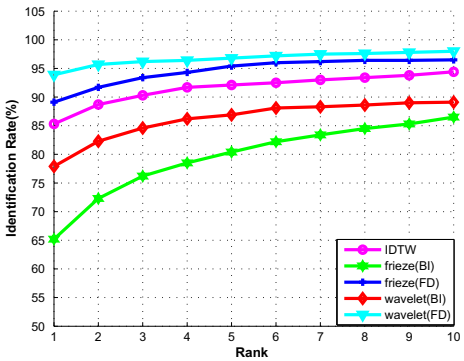

Fig. 9. Average CMS curves of all view angles.

Table 7
Performance comparison of the FDEI representation and the binary silhouette.

Methods	Angle	N_{ss}	N_{sf}	N_{fs}	N_{ff}	Z-value	Confidence (%)
Frieze(FD)	72	1179	601	20	60	23.0894	100
Vs.	90	1647	124	35	54	6.9789	100
Frieze(BI)	162	1425	246	52	137	11.1802	100
Wavelet(FD)	72	1514	284	17	45	15.3320	100
Vs.	90	1743	82	5	30	8.1481	100
Wavelet(BI)	162	1490	234	17	119	13.6338	100

5. Conclusion

In this letter, we propose a robust dynamic gait representation scheme, frame difference energy image (FDEI), to lower the influence of silhouette incompleteness. The gait shape aspects are embodied by the DEI, whereas the dynamic aspects are the positive portions of the frame difference. The proposed representation keeps most of the shape details and the gait temporal variation. Experimental results of both synthesized and real databases testified that the FDEI is a feasible gait representation. It gives better and more stable results than the binary image despite which feature is used. It also shows its superiority over other noise-insensitive algorithms, such as GEI, GHI, GMI and IDTW. The FDEI representation has a promising performance when occlusion or incompleteness occurs in the silhouettes.

Acknowledgement

This letter is partially supported by the NSFC(60402038 and 60872154), NBRPC(2006CB705700), the Chair Professors of the Cheung Kong Scholars, and the Program for Cheung Kong Scholars and Innovative Research Team in University (PCSIRT, Grant No. IRT0645). The authors would like to thank the associate editor and the anonymous reviewers for their constructive comments.

They would also like to thank the Karen von Deneen for her help on improving the manuscript.

References

Boulgouris, N.V., Hatzinakos, D., Plataniotis, K.N., 2005. Gait recognition: A challenging signal processing technology for biometric identification. IEEE Signal Process. Mag. 22, 78–90.

Chen, C., Liang, J., Zhao, H., Hu, H., 2006. Gait recognition using Hidden Markov model. In: Second Internat. Conf. on Natural Computation, LNCS, vol. 4221, part I, pp. 399–407.

Cheng, M.H., Ho, M.F., Huang, C.L., 2008. Gait analysis for human identification through manifold learning and HMM. Pattern Recognition 41 (8), 2541–2553.

Clark, A.F., Clark, C., 1999. Performance Characterization in Computer Vision: A Tutorial. <http://peipa.essex.ac.uk/benchmark/tutorials/essex/tutorial.pdf>.

Elgammal, A., Harwood, D., Davis, L., 2000. Non-parametric model for background subtraction. In: Sixth European Conf. on Computer Vision. Dublin, Ireland.

Friedman, N., Russell, S., 1997. Image segmentation in video sequences: A probabilistic approach. In: 13th Conf. on Uncertainty in Simulating Intelligence. Rhode Island, USA, pp. 175–181.

Gross, R., Shi, J., 2001. The Cmu Motion of Body (mobo) Database. Technical report, Robotics Institute.

Han, J., Bhanu, B., 2006. Individual recognition using Gait Energy Image. IEEE Trans. Pattern Anal. Machine Intell. 28 (2), 316–322.

Kale, A., Cuntoor, N., Chellappa, R., 2002. A framework for activity-specific human identification. In: Internat. Conf. on Acoustics, Speech and Signal Processing, pp. 706–714.

Little, J.J., Boyd, J.E., 1998. Recognizing people by their gait: The shape of motion. Videre 1, 1–32.

Liu, Z., Sarkar, S., 2005. Effect of silhouette quality and hard problems in gait recognition. IEEE Trans. Systems Man and Cybernetics. Part B: Cybernetics 35 (2), 170–183.

Liu, J., Zheng, N., 2007. Gait History Image: A novel temporal template for gait recognition. In: IEEE Internat. Conf. on Multimedia and Expo, pp. 663–666.

Liu, Y., Collins, R., Tsin, Y., 2002. Gait sequence analysis using frieze patterns. In: European Conf. on Computer Vision, pp. 659–671.

Liu, Z., Malave, L., Sarkar, S., 2004. Studies on silhouette quality and gait recognition. In IEEE Internat. Conf. on Computer Vision and Pattern Recognition, pp. 704–711.

Ma, Q., Wang, S., Nie, D., Qiu, J., 2007. Recognizing humans based on Gait Moment Image. In: Eighth ACIS Internat. Conf. on SNPD, pp. 606–610.

Nixon, M.S., Carter, J.N., 2006. Automatic recognition by gait. Proc. IEEE 94 (11), 2013–2024.

Sundaresan, A., RoyChowdhury, A., Chellappa, R., 2003. A hidden Markov Model Based Framework for Recognition of Humans from Gait Sequences. In: IEEE Internat. Conf. on Image Processing, pp. 93–96.

Tian, T.Y., Shah, M., 1996. Motion estimation and segmentation. Machine Vision Appl., 32–42.

Wang, L., Tan, T., Ning, H., Hu, W., 2004. Fusion of static and dynamic body biometrics for gait recognition. IEEE Trans. Circuits Systems Video Technol. 14 (2), 149–158.

Yacob, Y., Black, M.J., 1999. Parameterized modeling and recognition of activities. In: Sixth Internat. Conf. on Computer Vision. India, pp. 120–127.

Yu, S., Tan, D., Huang, K., Tan, T., 2007. Reducing the effect of noise on human contour in gait recognition. In: Internat. Conf. on Biometrics, pp. 338–346.