

# Gait Recognition Using Image Self-Similarity

**Chiraz BenAbdelkader**

*Identix Corporation, One Exchange Place, Jersey City, NJ 07302, USA*  
Email: [chiraz@cs.umd.edu](mailto:chiraz@cs.umd.edu)

**Ross G. Cutler**

*Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399, USA*  
Email: [rcutler@microsoft.com](mailto:rcutler@microsoft.com)

**Larry S. Davis**

*Department of Computer Science, University of Maryland, College Park, MD 20742, USA*  
Email: [lsd@umiacs.umd.edu](mailto:lsd@umiacs.umd.edu)

*Received 30 October 2002; Revised 18 May 2003*

Gait is one of the few biometrics that can be measured at a distance, and is hence useful for passive surveillance as well as biometric applications. Gait recognition research is still at its infancy, however, and we have yet to solve the fundamental issue of finding gait features which at once have sufficient discrimination power and can be extracted robustly and accurately from low-resolution video. This paper describes a novel gait recognition technique based on the image self-similarity of a walking person. We contend that the similarity plot encodes a projection of gait dynamics. It is also correspondence-free, robust to segmentation noise, and works well with low-resolution video. The method is tested on multiple data sets of varying sizes and degrees of difficulty. Performance is best for fronto-parallel viewpoints, whereby a recognition rate of 98% is achieved for a data set of 6 people, and 70% for a data set of 54 people.

**Keywords and phrases:** gait recognition, human identification at a distance, human movement analysis, behavioral biometrics, pattern recognition.

## 1. INTRODUCTION

### 1.1. Motivation

Gait is a relatively new and emergent behavioral biometric [1, 2] that pertains to the use of an individual's walking style (or "the way he walks") to determine identity. *Gait recognition* is the term typically used in the computer vision community to refer to the automatic extraction of visual cues that characterize the motion of a walking person in video and is used for identification purposes. Gait is particularly an attractive modality for passive surveillance since, unlike most biometrics, it can be measured at a distance, hence not requiring interaction with or cooperation of the subject. However, gait features exhibit a high degree of intraperson variability, being dependent on various physiological, psychological, and external factors such as footwear, clothing, surface of walking, mood, illness, fatigue, and so forth. The question then arises as to whether there is sufficient gait variability *between* people that can discriminate them even in the presence of large variation *within* each individual.

There is indeed strong evidence originating from psychophysical experiments [3, 4, 5] and gait analysis research

(a well-advanced multidisciplinary field that spans kinesiology, physiotherapy, orthopedic surgery, ergonomics, etc.) [6, 7, 8, 9, 10] that gait dynamics contain a signature that is characteristic of, and possibly unique to, each individual.

From a biomechanics standpoint, human gait consists of synchronized, integrated movements of hundreds of muscles and joints of the body. These movements follow the same basic bipedal pattern for all humans, and yet vary from one individual to another in certain details (such as their relative timing and magnitudes) as a function of their entire musculo-skeletal structure, that is, body mass, limb lengths, bone structure, and so forth. Because this structure is difficult to replicate, gait is believed to be unique to each individual and can be completely characterized by a few hundred kinematic parameters, namely, the angular velocities and accelerations at certain joints and body landmarks [6, 7]. Achieving such a complete characterization *automatically* from low-resolution video remains an open research problem in computer vision. The difficulty lies in that feature detection and tracking is error prone due to self-occlusions, insufficient texture, and so forth. This is why computer-aided motion analysis systems still rely on special

wearable instruments, such as LED markers, and walking surfaces [9].

Luckily, we may not need to recover 3D kinematics for gait recognition after all. In Johansson's early psychophysical experiments [3], human subjects were able to recognize the type of movement solely by observing light bulbs attached to a few joints of the moving person. The experiments were filmed in total darkness so that only the bulbs, a.k.a. moving light displays (MLDs), are visible. Similar experiments later suggested that the identity of a familiar person ("a friend") [5], as well as the gender of the person [4], may be recognizable from their MLDs. While it is widely agreed that these experiments provide evidence about motion perception in humans, there is no consensus on how the human visual system actually interprets this MLD-type stimuli. Two main theories exist: the first maintains that people recover the 3D structure of the moving object (person) and subsequently uses it for recognition; the second theory states that motion information is directly used for recognition, without structure recovery in the interim [11]. This seems to suggest that the raw spatiotemporal (XYT) patterns generated by the person's motion in an MLD video encode information that is sufficient to recognize their movement.

In this paper, we describe a novel gait recognition technique that derives classification features directly from these XYT patterns. Specifically, it computes the image self-similarity plot (SSP), defined as the correlation of all pairs of images in the sequence. Normalized feature vectors are extracted from the SSP and used for recognition. Related work has demonstrated the effective use of SSP's in recognizing different types of biological periodic motions, such as those of humans and dogs, and applied the technique for human detection in video [12]. We use them here to classify the movement patterns of different people. We contend that the SSP encodes a projection of planar gait dynamics and hence a 2D signature of gait. Whether it contains sufficient discriminant power for accurate recognition is what we set to determine.

As in any pattern recognition problem, these methods typically consist of two stages: a feature extraction stage that derives motion information from the image sequence and organizes it into some compact form (or representation), and a recognition stage that applies some standard pattern classification technique to the obtained motion patterns, such as K-nearest neighbor (KNN), support vector machines (SVM), and hidden Markov models (HMM). In our view, the crux of the gait recognition problem lies in perfecting the first stage. The challenge is in finding motion patterns that are sufficiently discriminant despite the wide range of natural variability of gait, and that can be extracted reliably and consistently from video. The method of this paper is designed with these two requirements in mind. It is based on the SSP which is robust to segmentation noise and can be computed correspondence-free from fairly low-resolution images. Although this method is view-dependent (since it is inherently appearance-based), this is circumvented via view-based recognition. The method is evaluated on several data sets of varying degrees of difficulty, including a large surveillance-

quality outdoor data set of 54 people, and a multiview data set of 12 people taken from 8 viewpoints.

## 1.2. Assumptions

The method makes the following assumptions:

- (i) people walk with constant velocity for about 3–4 seconds;
- (ii) people are located sufficiently far from the camera;
- (iii) the frame rate is greater than twice the frequency of the walking;
- (iv) the camera is stationary.

## 1.3. Organization of the paper

The rest of the paper is organized as follows. Section 2 discusses related work in the computer vision literature and Section 3 describes the method in detail. We assess the performance of the method on a number of different data sets in Section 4, and finally conclude in Section 5.

## 2. RELATED WORK

Interest in gait recognition is best evidenced by the near-exponential growth of the size of related literature over the past few years [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32]. Gait recognition is generally related to human movement analysis methods that automatically detect and/or track human motion in video for a variety of applications-surveillance, videoconferencing, man-machine interfaces, smart rooms, and so forth. For good surveys on this topic, see [11, 33, 34]. It is perhaps most closely associated with the subset of methods that analyze whole-body movement, for example, for human detection [12, 35, 36] and activity recognition [37, 38, 39, 40].

A common characteristic to all of these methods is that they consist of two main stages: a feature extraction stage in which motion information is derived from the image sequence and organized into some compact form (or representation), and a recognition stage in which some standard pattern classification technique is applied to the obtained motion patterns, such as KNN, SVM, and HMM. We distinguish two main classes of gait recognition methods: holistic [14, 15, 16, 17, 18, 19, 23, 24, 25, 28, 29, 30, 31, 32] and feature-based [20, 21, 22, 26, 27, 41, 42, 43, 44]. The holistic versus feature-based dichotomy can also be regarded as global versus local, nonparametric versus parametric, and pixel-based versus geometric. This dichotomy is certainly recurrent in pattern recognition problems such as face recognition [45, 46]. In the sequel, we describe and critique examples from both approaches, and relate them to our gait recognition method.

### 2.1. Holistic approach

The holistic approach characterizes body movement by the statistics of the XYT patterns generated in the image sequence by the walking person. Although typically these patterns have no direct physical meaning, intuitively they capture both the static and dynamic properties of body shape. There are many ways of extracting XYT patterns from the

image sequence of a walking person. However, in a nutshell, they all either extract raw XYT data (namely, the temporal sequence of binary/color silhouettes or optical flow images), or a mapping of this data to a more terse 1D or 2D signal.

Perhaps the simplest approach is to use the sequence of binary silhouettes spanning one gait cycle and scaled to a certain uniform size [15, 32]. The method of [30] differs slightly from this in that it uses silhouettes corresponding to certain gait poses only, namely, the double-support and mid-stance poses. Classification is achieved either by directly comparing (correlating) these silhouette sequences [30, 32] or by first projecting them onto a smaller subspace (using principal components analysis [15] and/or Fisher's linear discriminant analysis [17]), then comparing them in this subspace. Although excellent classification rates are reported for some of these methods (particularly [30]), they are the most sensitive (among holistic methods) to any variation in the appearance of the silhouette whether due to clothing and camera viewpoint or to segmentation noise. Nonetheless, these methods are the simplest and hence provide good baseline performance against which to evaluate other more contrived gait recognition methods.

Rather than using the entire silhouette, other methods use a signature of the silhouette by collapsing the XYT data into a more terse 1D or 2D signal(s), such as binary shape moments, vertical projection histograms (XT), and horizontal projection histograms (YT) [14, 18, 28, 30, 31]. Niyogi and Adelson [14] extract four (2D) XT sheets that encode the person's inner and outer bounding contours. Similarly, Liu et al. [31] extract the XT and YT projections of the binary silhouettes. He and Debrunner [18] compute a quantized vector of Hu moments from the person's binary silhouette at discrete gait poses and use them for recognition via an HMM. The method of Kale et al. [28] is quite similar to this, except that they use the vector of silhouette widths (for each latitude) instead of Hu moments. Certainly, the SSP of the present paper is a mapping of the sequence of silhouettes to a 2D signal. However, while the SSP is quite robust to the segmentation noise in binary silhouettes, signals derived directly from binary silhouettes are typically very sensitive to segmentation noise even with smoothing.

A third category of holistic methods apply two levels of aggregation on the XYT data, and not one [16, 19, 23, 29]. They first map the XYT data of the walking person into one or more 1D signals, then aggregate these into a feature vector by computing the statistics of these signals (such as their first- and second-order moments). Lee and Grimson [29] fit ellipses to seven rectangular subdivisions of the silhouette then compute four statistics (first and second-order moments) for each ellipse, and hence obtain 28 1D signals from the entire silhouette sequence. Finally, they use three different methods for mapping these signals to obtain a single feature vector for classification.

Little and Boyd [16] use optical flow instead of binary silhouettes. They fit an ellipse to the dense optical flow of the person's motion, then compute thirteen scalars consisting of first- and second-order moments of this ellipse. Periodicity analysis is applied to the resulting thirteen 1D sig-

nals, and a 12D feature vector is computed consisting of the phase difference between one signal and all other twelve signals. Recognition is achieved via exemplar KNN classification in this 12D feature space. These features are both scale-invariant and time-shift invariant so that no temporal scaling nor alignment is necessary.

Obviously, the advantage of the holistic approach lies in that it is correspondence-free, and hence simple to implement. Its main drawback is that the extracted features are inherently appearance-based, and hence likely to be sensitive to any factors that alter the person's silhouette, particularly camera viewpoint and clothing. Viewpoint dependence can be remedied by estimating the viewpoint of the walking person and using view-based recognition. However, it is not obvious how or whether the clothing sensitivity problem could be solved.

## 2.2. Feature-based approach

The feature-based approach recovers explicit features (or parameters) describing gait dynamics, such as stride dimensions and the kinematics, of joint angles. Although human body measurements (i.e., absolute distances between certain landmarks, such as height, limb lengths, shoulder width, head circumference, etc.) are not descriptors of body movement, they are indeed determinants of that movement, and hence can also be considered as gait parameters.

Bobick and Johnson [22] compute body height, torso length, leg length, and step length for identification. Using a priori knowledge about body structure at the double-support phase of walking (i.e., when the feet are maximally apart), they estimate these features as distances between fiducial points (namely, the midpoint and extrema) of the binary silhouette. Obviously, the accuracy of these measurements is very sensitive to segmentation noise in the silhouette, even if they are averaged over many frames.

In [42], Davis uses a similar approach to compute the stride length and cadence, though he relies on reflective markers to track 3D trajectories of head and ankle. With measurements obtained from 12 people, he is able to train a linear perceptron to discriminate the gaits of adults and children (3–5 years old) to within 93% accuracy. BenAbdelkader et al. describe a more robust method to compute stride dimensions, which exploits not only the periodicity of walking but also the fact that people walk in contiguous steps [44]. In related work [26], they further estimate the height variation of a walking person by fitting it to a sinusoidal model and use the two model parameters along with the stride dimensions for identification.

The kinematics of a sufficient number of body landmarks can potentially provide a much richer, and perhaps unique, description of gait. Bissacco et al. [27] fit the trajectories of 3D joint positions and joint angles to a discrete-time continuous-state dynamical system. They use the space spanned by the parameters of this model for recognizing different gaits. Tsai et al. [41] use one cycle of the XYZ curvature function of 3D trajectories of certain points on the body for identification.

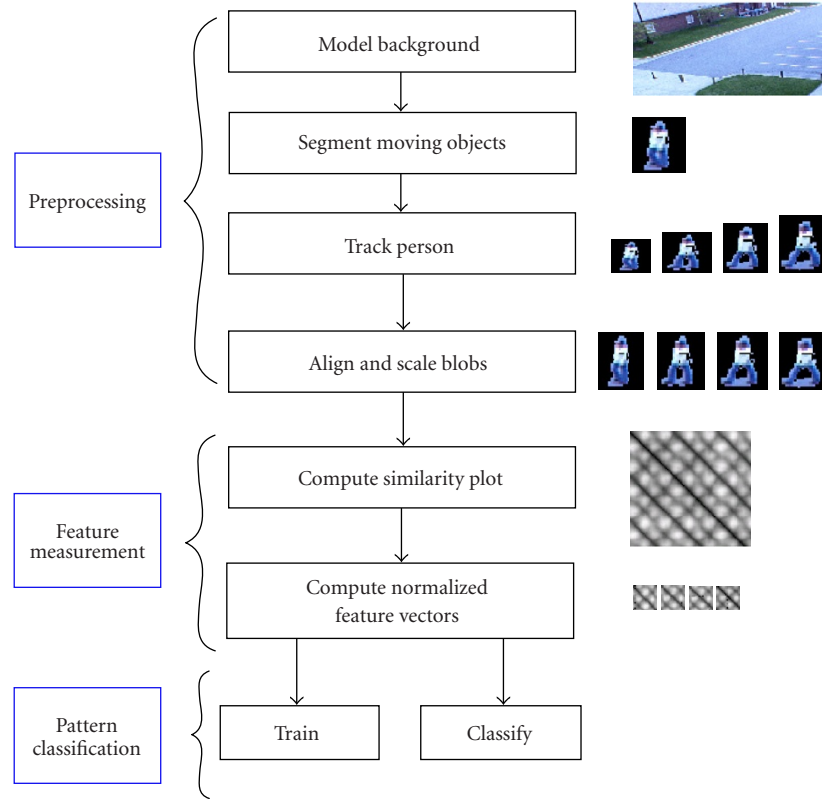


FIGURE 1: Overview of method.

The major strength of this approach lies in that it uses classification features that are known to be directly pertinent to gait dynamics, unlike its holistic counterpart. Another advantage is that it is in principle view-invariant since it uses 3D quantities for classification. However, its measurement accuracy degrades for certain viewpoints as well as at low resolutions. Obviously, accurate measurement of most of these gait parameters requires not only accurate camera calibration but also accurate detection and tracking of anatomical landmarks in the image sequence. The feasibility of this approach is currently very limited mainly due to the difficulty of automatic detection and tracking in realistic (low-resolution) video. For example, all of [27, 41, 42] use 3D motion capture data or semimanually tracked features in order to avoid the automatic detection and tracking problem altogether.

### 3. METHOD

The proposed gait recognition method characterizes gait in terms of a 2D signature computed directly from the sequence of silhouettes, that is, the XYT volume of the walking person. This signature consists of the SSP, which was first introduced in [47] for the purpose of motion classification, and is defined as the matrix of cross-correlation between each pair of images in the sequence. The SSP has the advantage of being correspondence-free and robust to segmentation and track-

ing errors. Also, intuitively, it can be seen that the SSP encodes both the static (first-order) properties and temporal variations of body shape during the walking.

The method can be seen as a generic pattern classifier [48, 49] composed of the three main modules shown in Figure 1. First, the moving person is segmented and tracked in each frame of the given image sequence (*preprocessing module*). Then the SSP is computed from the obtained silhouette sequence, and properly aligned and scaled to account for differences in gait frequency and phase, thus obtaining a set of normalized feature vectors (*feature measurement module*). Finally, the person's identity is determined by applying standard classification techniques on the normalized feature vectors (*pattern classification module*). Sections 3.1, 3.2, and 3.3 discuss each of these modules in detail.

#### 3.1. Preprocessing

Given a sequence of images obtained from a static camera, we detect and track the moving person then compute the corresponding sequence of motion regions (or blobs) in each frame. Motion segmentation is achieved via a nonparametric background modeling/subtraction technique that is quite robust to lighting changes, camera jitter, and to the presence of shadows [50]. Once detected, the person is tracked in subsequent frames via simple spatial coherence, namely based on the overlap of blob bounding boxes in any two





FIGURE 2: The SSP can be computed from the sequence of silhouettes corresponding to the original image, the foreground image, or the binary image (from left to right).

consecutive frames [51]. The issue of determining whether a foreground blob indeed corresponds to a moving person is addressed in the feature measurement module.<sup>1</sup> Specifically, we use the cadence-based technique described in [35] which simply verifies whether the computed cadence is within the normal range of human walking (roughly 80–145 steps/m).

Once a person has been tracked for  $N$  consecutive frames, a sequence of  $N$  corresponding silhouette templates is created as follows. Given the person's blob in each frame, we extract the (rectangular) region<sup>2</sup> enclosed within its bounding box either from (1) the original color/greyscale image, (2) the foreground image, or (3) the binary image, as shown in Figure 2. Clearly, there are competing trade-offs to using either type of template in measuring image similarity (when computing the SSP). The first is more robust to segmentation errors. The third is more robust to clothing and background variations. The second is simply a hybrid of these two; it is robust to background variations but sensitive to segmentation errors and clothing variations.

### 3.2. Feature measurement

#### 3.2.1. Silhouette template scaling

The silhouette templates need to be first scaled to a standard size to normalize for depth variations (Figure 3). It is worth noting that this will only work for small depth changes. Large depth changes may introduce nonlinear variations (such as loss of detail and perspective effects) and hence cannot be normalized merely via a linear scaling of the silhouettes.

The apparent size of a walking person varies at the frequency of gait, due to the pendular-like oscillatory motion of the legs and arms, and consequently the width and height of a person's image also vary at the fundamental frequency of walking. Specifically, let  $w(n)$  and  $h(n)$  be the width and

height of the  $n$ th image (template) of the person. According to gait analysis literature [6],  $w(n)$  and  $h(n)$  can be approximated as sinusoidal functions:

$$\begin{aligned} w(n) &= m_w(n) + A_w \sin \omega n + \phi, \\ h(n) &= m_h(n) + A_h \sin \omega n + \phi, \end{aligned} \quad (1)$$

where  $\omega$  is the frequency of gait (in radians per frame) and  $\phi$  is the phase of gait (in radians). Note that  $m_w(n)$  is the mean width and  $A_w$  is the amplitude of oscillation (around this mean). The same can be said about  $m_h(n)$  and  $A_h$ , respectively, for height. Furthermore, in fronto-parallel walking,  $m_w(n)$  and  $m_h(n)$  are almost constant, while in non-fronto-parallel walking, and due to the changing camera depth, they increase/decrease approximately linearly (i.e., in a linear trend):  $m_w(n) \cong \alpha_w n + \beta_w$  and  $m_h(n) \cong \alpha_h n + \beta_h$ . Figure 3 illustrates these two different cases.

Therefore, in order to account for template size variation caused by camera depth changes (during non-fronto parallel walking), we first de-trend them:

$$\begin{aligned} \hat{w}(n) &= w(n) - \alpha_w n = \beta_w + A_w \sin \omega n + \phi, \\ \hat{h}(n) &= h(n) - \alpha_h n = \beta_h + A_h \sin \omega n + \phi, \end{aligned} \quad (2)$$

so that the templates now have equal mean width and height. Note, however, that we need  $\hat{w}(n)/w(n) = \hat{h}(n)/h(n)$  for all  $n$ , that is,  $\alpha_w/\alpha_h = w(n)/h(n)$ , so that each template can be uniformly scaled along its width and height. In other words, we need the width-to-height aspect ratio to remain constant throughout the sequence. This is a valid assumption since the person is sufficiently far from the camera, barring abrupt/sharp changes in person's pose with respect to the camera.

Finally, the templates are scaled one more time so that their mean height is equal to some given constant  $H_0$  (we typically use  $H_0 = 50$  pixels):

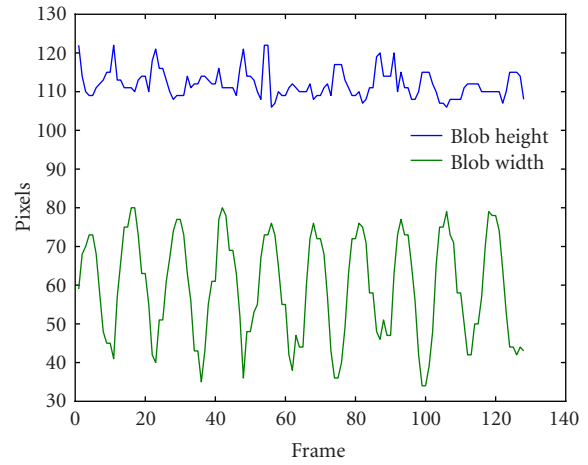
$$\tilde{h}(n) = \hat{h}(n) \cdot \frac{H_0}{\beta_h} = H_0 + \tilde{A}_h \sin \omega n + \phi. \quad (3)$$

<sup>1</sup>The only reason this is *not* done in the current module is for the sake of modularity, since cadence is computed in the second module.

<sup>2</sup>The cropped region also includes an empty 10-pixel border in order to allow for shifting when we later compute the cross-correlation of template pairs.



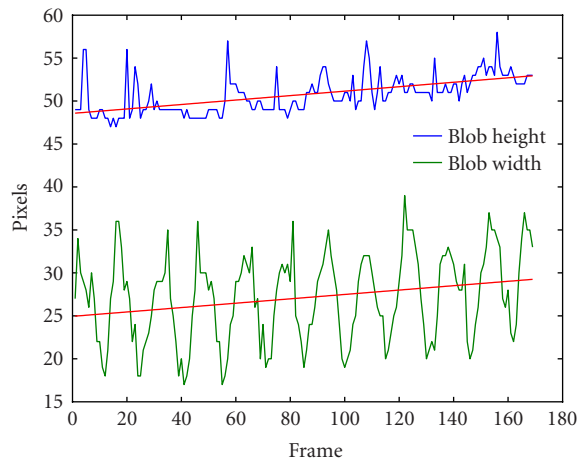
(a)



(b)



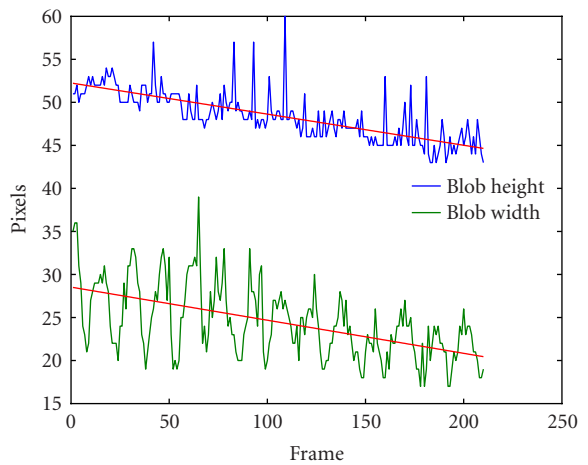
(c)



(d)



(e)



(f)

FIGURE 3: Template dimensions in pixels for (a), (b) a fronto-parallel sequence, (c), (d), (e), and (f) two non-fronto-parallel sequences (bottom two rows). The width and height increase when the person walks closer to the camera (middle row), and decrease as the person moves away from the camera (bottom row). The red lines correspond to the linear trend in both these cases.

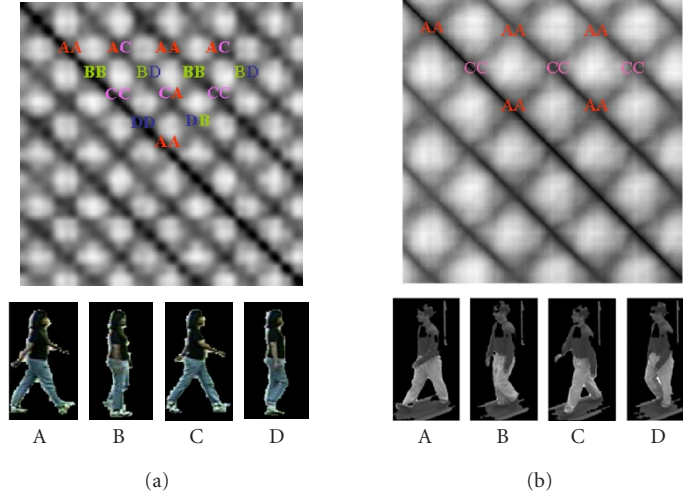


FIGURE 4: The SSPs for (a) a fronto-parallel sequence and (b) a non-fronto-parallel sequence computed here using foreground templates. Similarity values are linearly scaled to the gray scale intensity range  $[0, 255]$  for visualization. The local minima of each SSP correspond to combinations of key poses of gait (labelled A, B, C, and D).

### 3.2.2. Computing the self-similarity plot

Let  $I_i$  be the  $i$ th scaled template with size  $\tilde{w}_i \times \tilde{h}_i$  (in pixels). The corresponding SSP  $S(i, j)$  is computed as the absolute correlation<sup>3</sup> of each pair of templates  $I_i$  and  $I_j$ , minimized over a small search radius  $r$ , namely,

$$S(i, j) = \min_{|dx| < r, |dy| < r} \sum_{|x| \leq W/2} \sum_{|y| \leq H/2} |I_j(x + dx, y + dy) - I_i(x, y)|, \quad (4)$$

where  $W = \min(\tilde{w}_i, \tilde{w}_j - 2r)$  and  $H = \min(\tilde{h}_i, \tilde{h}_j - 2r)$  so that the summation does not go out of bounds. Although ideally  $S$  should be symmetric, it typically is not, unless  $r = 0$ .

Figure 4 highlights some of the properties of  $S$  for fronto-parallel and non-fronto-parallel walking sequences. The diagonals are due to the periodicity of gait, while the cross-diagonals are due to the temporal mirror symmetry of the gait cycle [47]. The intersections of these diagonals, that is, the local minima of  $S$ , correspond to key poses of the gait cycle: the mid-stance (B and D) and double-support (A and C) poses. Thus  $S$  encodes both the frequency and phase of the gait cycle. Some of these intersections disappear for non-fronto-parallel sequences (BD, BB, and DD) because gait does not appear bilaterally symmetric.

### 3.2.3. Normalizing the self-similarity plot

Since we are interested in using the SSP for recognition, we need to be able to compare the SSPs of two different walk-

ing sequences. Furthermore, gait consists of repeated steps, and so it only makes sense to compare two SSPs that contain an *equal* number of walking cycles and start at the *same* phase (i.e., body pose). In other words, we need to normalize the SSP for differences in sequence length and starting phase. There are several ways to achieve this. In a previous work, we used a submatrix of the SSP that starts at the first occurrence of the double-support pose<sup>4</sup> in the sequence and spans three gait cycles (i.e., six steps) [52].

A different approach that proves to be better for recognition [25] uses the so-called self-similarity units (SSUs). Each SSU is a submatrix of the SSP that starts at the double-support phase and spans one gait cycle. The SSP can then be viewed as a tiling of (contiguous) SSUs, and a different tiling can be obtained for any particular starting phase. We use all SSUs corresponding to the left and right double-support poses for gait recognition. However, because the SSP is (approximately) symmetric and for computational efficiency, we only use the SSUs of the top half, as shown in Figure 5. We can easily show that for a sequence containing  $K$  gait cycles, there are  $2(K(K+1)/2) = K(K+1)$  SSUs.

Finally, because the size of each SSU is defined both by the duration of a gait cycle and the frame rate (namely,  $P = T \cdot F_s$  frames, where  $T$  is the average gait cycle length in seconds and  $F_s$  is the frame rate), we scale all SSUs to some uniform size of  $m \times m$  in order to be able to compare them.

### 3.2.4. Computing the frequency and phase of gait

Obviously, we need to compute the frequency and phase of gait in order to normalize the SSP and obtain the SSUs.

<sup>3</sup>We chose absolute correlation for its simplicity. Other similarity measures include normalized cross-correlation, the ratio of overlapping foreground pixels, Hausdorff distance, and so forth.

<sup>4</sup>The double-support phase of the gait cycle corresponds to when the feet are maximally apart. The left double-support pose is when the left leg is leading and the right double-support pose is when the right leg is leading.



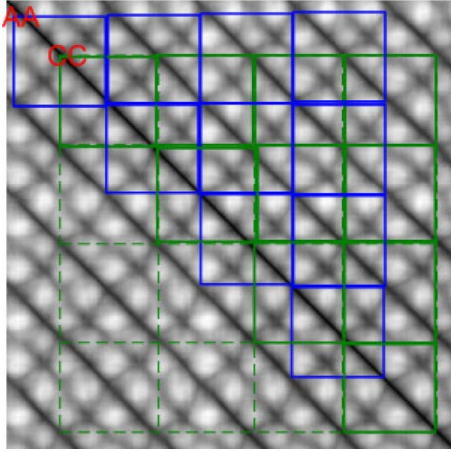


FIGURE 5: Extracting SSUs from the similarity plot. Blue and green SSUs start at pose A and C, respectively.

Several methods in the vision literature have addressed this problem, typically via periodicity analysis of some feature of body shape or texture [12, 53, 54]. In fact, most existing gait recognition methods involve some type of frequency/phase normalization, and hence devise some method for computing the frequency and phase of gait.

In this paper, we compute gait frequency and phase via analysis of the SSP, which indeed encodes the frequency and phase of walking, as mentioned in Section 3.2.2. We found this to be more robust than using, say, the width or height of the silhouette, as we have done in the past [52]. For the frequency, we apply the autocorrelation method on the SSP as was done in [12]. This method is known to be more robust to nonwhite noise and nonlinear amplitude modulations than Fourier analysis. It first smoothes the autocorrelation matrix of the SSP, computes its peaks, then finds the best-fitting regular 2D lattice for these peaks. The period is then obtained as the width of this best-fitting lattice.

The phase is computed by locating the local minima of the SSP that correspond to the A and C poses (defined in Section 3.2.2). However, not *all* local minima correspond to these two poses, since in near-fronto-parallel sequences, combinations of the B and D poses also form a local minima. Fortunately, the two types of local minima can be distinguished by the fact that those corresponding to A and C poses are “flatter” than those corresponding to B and D poses. However, we are still only able to resolve the phase of gait up to half a period, since we have no way of distinguishing the A and C poses from one another. As a result, the SSUs corresponding to both A and C poses (shown in Figure 5) are all used for gait recognition.

### 3.3. Pattern classification

We formulate the problem as one of supervised pattern classification. Given a labeled set of SSUs (wherein each SSU has the label of the person it corresponds to), termed the *gallery*, we want to determine the person corresponding to a set of novel (unknown) SSUs, termed the *probe*. This can

be achieved in two steps: (1) pattern matching, which computes some measure of the degree of match (or mismatch) between each pair of probe and gallery patterns and (2) decision, which determines the probe’s correct class based on these match (or mismatch) scores. For the latter, we simply use a variation of the KNN rule. For the former, we use two different approaches, namely, template matching (TM) and statistical pattern classification, discussed separately in Sections 3.3.1 and 3.3.2.

#### 3.3.1. Template matching

Because the SSU is an  $m \times m$  2D template, perhaps the simplest distance metric between two SSUs is their maximum cross-correlation computed over a small range of 2D shifts (we typically use the range  $[-5, 5]$ ). The advantage of this approach is that it explicitly compensates for small phase alignment errors. Its disadvantage is that it is computationally very demanding.

#### 3.3.2. Statistical pattern classification

Here, each SSU is represented as a  $p$ -dimensional vector,  $p = m^2$ , by concatenating its  $m$  rows. The distance between two patterns is then simply computed as their Euclidean distance in this space. However, when  $p$  is large, it is desirable to first reduce the dimensionality of the vector space for the sake of computational efficiency as well as to circumvent the curse of dimensionality phenomenon [48, 49, 55].

Dimensionality reduction, also called feature extraction, maps the vectors to a  $q$ -dimensional space with  $q \ll p$ . We consider three linear feature extraction techniques for this problem: principal component analysis (PCA), linear discriminant analysis (LDA), and a so-called subspace-LDA (s-LDA) that combines the latter two techniques by applying LDA on a subspace spanned by the first few principal components. See [56, 57, 58, 59, 60, 61] for examples of the application of these methods in face recognition.

Each method defines a linear transformation  $W$  that maps a  $p$ -dimensional vector  $u$  in the original feature space onto a  $q$ -dimensional vector  $\zeta = (\zeta_1, \dots, \zeta_q)$  such that  $\zeta = W^T u$ . Note that  $(\zeta_1, \dots, \zeta_q)$  can also be viewed as the coordinates of  $u$  in this  $q$ -dimensional subspace.

The  $p \times q$  matrix  $W$  is determined from a given training set of vectors by optimizing some objective criterion. The choice of  $q$  seems to be domain-dependent and we have not as yet devised a method to automatically select it. Instead, we simply choose the value that achieves best classification rate for the given training and test data sets.

Choosing between PCA, LDA, and s-LDA is also domain-dependent. It depends on the relative magnitudes of the within-class scatter and the between-class scatter, as well as the size of the training set. Furthermore, one design issue common to all three approaches is the choice of the subspace dimensionality.

## 4. EXPERIMENTS AND RESULTS

We evaluate the performance of the method on four different data sets of varying degrees of difficulty, and use the holdout



(also called split-sample) cross-validation technique to estimate the classification error rate for each data set [55]. Our goal is to quantify the effect of the following factors on performance.

- (i) *Natural individual variability* due to various physical and psychological factors such as clothing, footwear, cadence, mood, fatigue, and so forth. This within-person variation is introduced by using multiple samples of each person's walking taken at different times and/or over different days. It is worth noting, however, that sequences taken on different days will typically contain unwanted variations such as background, lighting, and clothing variations, which makes the recognition task even more difficult.
- (ii) *Photometric parameters*, namely, camera viewpoint, camera depth, and frame sampling rate.
- (iii) *Algorithm design parameters*, namely, the image similarity metric (correlation of binary silhouettes (BC) and correlation of foreground silhouettes (FC)), the pattern matching approach (PCA, LDA, s-LDA, and TM), and the KNN classifier parameter ( $K = 1, 3$ ).

#### 4.1. Data set 1

This data set is the same used by Little and Boyd in [16]. It consists of 42 image sequences with six different subjects (4 males and two females), 7 sequences of each, taken from a static camera at 30 fps and  $320 \times 240$  resolution. The subjects walked a fixed path against a uniform background. Thus the only source of variation in this data set (aside from random measurement noise) is the individuals' own walking variability across different samples.

Figure 6 shows all seven subjects overlaid on the background image. Figure 7 shows three of the SSP's for each person in Figure 6. The results are shown in Table 1. Note that LDA is not used for this data set because the number of training samples is insufficient for this kind of analysis [48]. Obviously, BC gives slightly better results than FC, and that s-LDA also slightly outperformed PCA. However, there is a significant improvement when using feature extraction (PCA and s-LDA) over TM.

#### 4.2. Data set 2

The second data set contains fronto-parallel sequences of 44 different subjects (10 females and 34 males), taken in an outdoor environment from two different cameras simultaneously, as shown in Figure 8. The two cameras are both fronto-parallel but located at different depths (approximately 20 ft and 70 ft) with respect to the walking plane. Each subject walked in two different sessions a fixed straight path, back and forth, at his/her natural pace. The sequences were captured at 20 fps and at full-color resolution of  $644 \times 484$ .

Six holdout experiments are carried out on this data set, with absolute correlation of BC used as the image similarity measure. The results are summarized in Table 2. The classification performance is better for the far camera (first row) than for the near camera (second row), which may be due to superior image quality of the far camera. Also, perfor-



FIGURE 6: The six subjects for data set 1, shown overlaid on the background image.

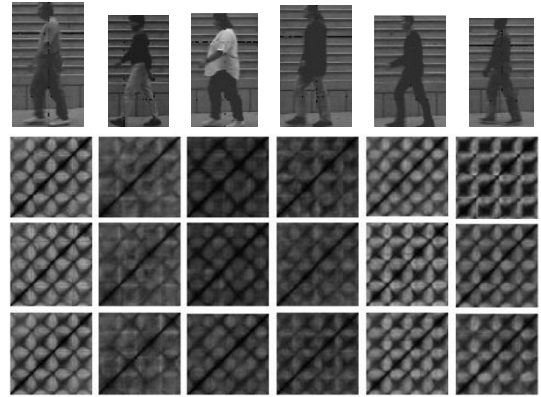


FIGURE 7: Three of the SSPs for each person in data set 1.

TABLE 1: Classification rates for the first data set for different image similarity metrics (BC and FC), pattern matching approaches (PCA, s-LDA, and TM), and KNN classifier parameters ( $K$ ).

$K$	BC			FC		
	PCA	s-LDA	TM	PCA	s-LDA	TM
1	94	100	89	94	94	94
3	94	100	94	94	94	94

mance degrades significantly when the training and test sets are from different cameras (third and fourth rows), which may be because our method is not invariant to large changes of camera depth, and hence confirms our observation in Section 3.2.1.

#### 4.3. Data set 3

In order to evaluate the performance of the method across large changes in camera viewpoint, we used the Keck multiperspective lab [62] to capture sequences of people walking on a treadmill from 8 different cameras at a time, as illustrated in Figure 9. The cameras are placed at the same height around half a circle so that they have the same tilt angle and different pan angles. The latter span a range of about 135 deg of the viewing sphere, though not uniformly. The data set contains 12 people (3 females and 9 males) and about 5

FIGURE 8: Second outdoor data set. Sample frames from (a) the *near* camera and (b) the *far* camera.

TABLE 2: Classification performance on the second data set using holdout technique with six different training and testing subsets.

Training set	Test set	PCA		LDA		s-LDA		TM
		$K = 1$	$K = 3$	$K = 1$	$K = 3$	$K = 1$	$K = 3$	$K = 1$
Far camera	Far camera	49	49	63	70	65	67	59
Near camera	Near camera	53	52	41	46	49	52	55
Far camera	Near camera	10	10	22	23	24	25	21
Near camera	Far camera	17	23	24	22	23	23	25
Both cameras	Both cameras	52	50	52	52	54	56	35

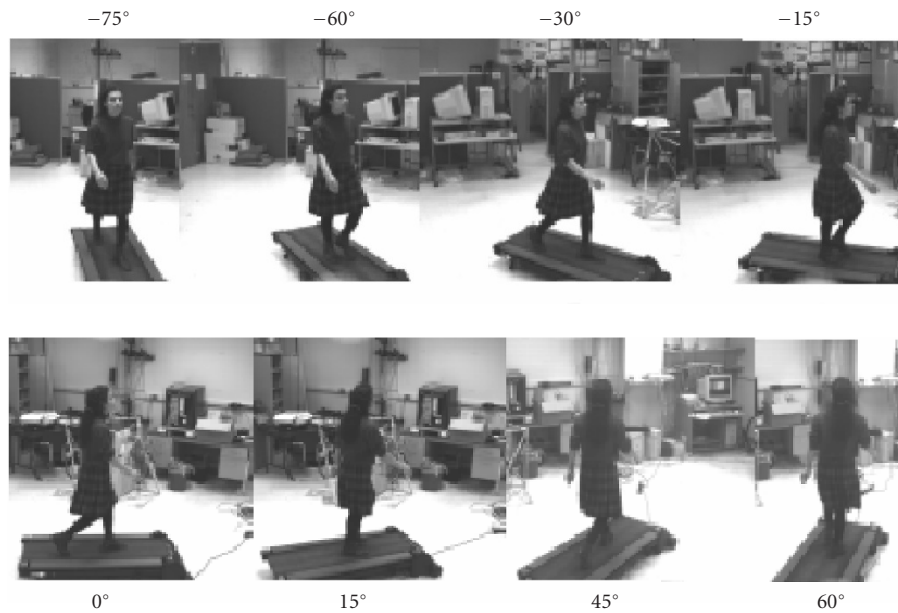


FIGURE 9: Eight camera viewpoints of the sequences in second test data set.

sequences per person per view on average, taken mostly on different days for each person. The sequences were captured at a frame rate of 60 fps and a resolution of  $644 \times 488$  greyscale images.

Like in general object recognition problems, there are two main approaches to gait recognition under variable viewing conditions: a view-based approach and a parametric approach. In the view-based approach, a classifier is trained

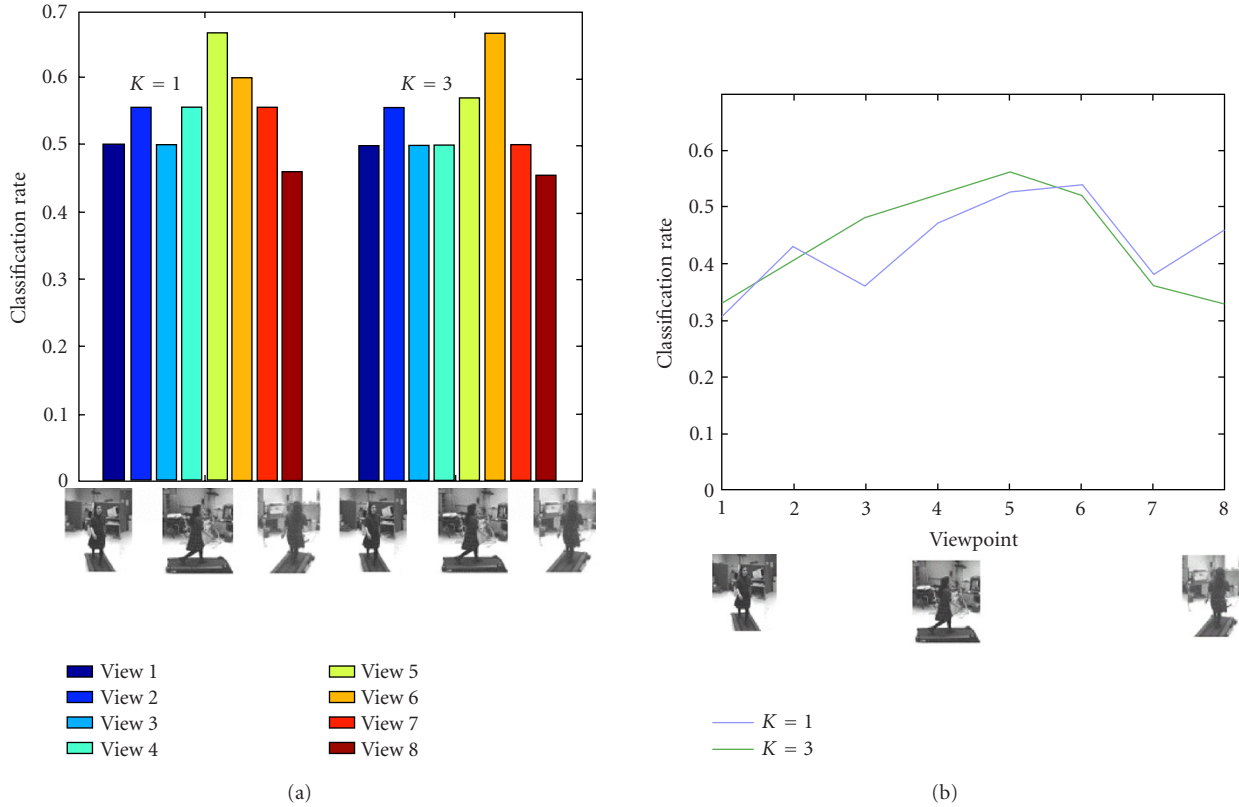


FIGURE 10: Classification performance for data set 3: (a) view-based approach, (b) parametric approach.

separately for each viewpoint, that is, there are as many classifiers as there are camera viewpoints. A novel sequence needs to first have its viewpoint determined so that the corresponding classifier is applied. The parametric approach, on the other hand, trains a single classifier using data from all viewpoints.

Both these approaches are applied to the data set and the results are shown in Figure 10. We use absolute correlation of binary silhouettes for image similarity, and the hold-out cross-validation technique to estimate the classification rate, whereby we train on data from six days and test on data from the seventh day. This is repeated 7 times, and the classification rate is computed as the average over the seven iterations. Clearly, the performance is best for near-fronto-parallel views (4–6). An intuitive explanation for this is the following. Most of the dynamics of walking takes place in the sagittal plane, which is the plane containing both legs. Hence in a non-fronto-parallel viewpoint, where the sagittal plane is almost orthogonal to the image plane, much less of the appearance variation caused by gait dynamics is captured, which may be insufficient for recognition. Furthermore, view-based approach gives overall better results than the parametric approach.

#### 4.4. Data set 4

Recall that we scale the SSUs to a fixed size ( $m \times m$  pixels), which is equivalent to normalizing the gait frequency

to a fixed value (i.e., temporal scaling). However, because gait dynamics are inherently a function of cadence, we expect that the SSUs corresponding to significantly different cadences to be *qualitatively* different (even if they are pre-normalized to the same frequency). We tested this expectation using a portion of CMU's MoBo data set [63], consisting of indoor sequences of 25 people walking on a treadmill and captured from 3 different views, as shown in Figure 11. Furthermore, each person walked at two different speeds: a slow pace (2.06 miles/h) and a moderate pace (2.82 miles/h). Thus we used a total of 150 sequences for this experiment (i.e., 2 sequences per person per view). The sequences are all captured on the same day and against the same background.

We then setup experiments via the holdout cross-validation technique in which the train and test sets correspond to different combinations of speeds. This was done for each view separately, however (i.e., view-based classification). Table 3 shows the classification results. As expected, the performance degrades significantly when the training and test data for the classifier correspond to different speeds.

## 5. CONCLUSIONS

We described a novel holistic gait recognition approach that uses image self-similarity as the basic feature for classification. The method is correspondence-free, works well with

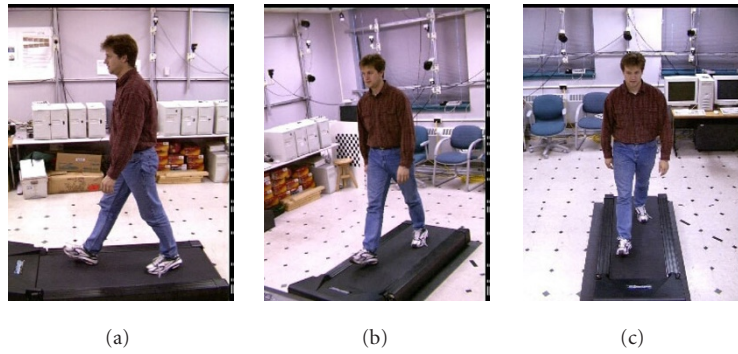


FIGURE 11: The three views in data set 4.

TABLE 3: Performance on the fifth data set: classification rates using holdout technique with four different training and testing subsets.

	Slow/slow	Fast/fast	Slow/fast	Fast/slow
View 1	100	100	54	32
View 2	100	96	26	16
View 3	96	100	43	33

low-resolution video, and is robust to variation in clothing, lighting, and to segmentation errors. A recognition rate of 100% is achieved for a fronto-parallel data set of 6 people, and 70% for a fronto-parallel data set of 54 people.

Although the method is inherently appearance-based, and hence view-dependent, this is circumvented via view-based recognition. Using a data set of 12 people captured from 8 viewpoints, recognition rates decreases from about 65% for near fronto-parallel viewpoints to about 47% for near frontal viewpoints. Performance also degrades when camera depth and cadence are significantly changed.

We are working to combine the gait features of this method with geometric gait features that can be robustly computed from video, such as cadence, stride length, and stature. We also plan to study the use of these features for other recognition tasks, such as gender classification and gait asymmetry detection caused, for example, by a limp.

## ACKNOWLEDGMENT

This paper was written under the support of DARPA's Human ID at a Distance Project.

## REFERENCES

- [1] A. Jain, *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Publishers, Boston, Mass, USA, 1999.
- [2] D. D. Zhang, *Automated Biometrics: Technologies and Systems*, Kluwer Academic Publishers, Boston, Mass, USA, 2000.
- [3] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception and Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.
- [4] J. E. Cutting and L. T. Kozlowski, "Recognizing friends by their walk: gait perception without familiarity cues," *Bulletin Psychonomic Soc.*, vol. 9, no. 5, pp. 353–356, 1977.
- [5] C. Barclay, J. E. Cutting, and L. T. Kozlowski, "Temporal and spatial factors in gait perception that influence gender recognition," *Perception and Psychophysics*, vol. 23, no. 2, pp. 145–152, 1978.
- [6] M. Murray, "Gait as a total pattern of movement," *American Journal of Physical Medicine*, vol. 46, no. 1, pp. 290–332, 1967.
- [7] H. J. Ralston, V. Inman, and F. Todd, *Human Walking*, Williams and Wilkins, Baltimore, MD, USA, 1981.
- [8] D. Winter, *The Biomechanics and Motor Control of Human Gait*, University of Waterloo Press, Waterloo, Ontario, Canada, 1987.
- [9] J. Perry, *Gait Analysis: Normal and Pathological Function*, Slack, Thorofare, NJ, USA, 1992.
- [10] J. Rose and J. G. Gamble, *Human Walking*, Williams and Wilkins, Baltimore, MD, USA, 2nd edition, 1994.
- [11] C. Cedras and M. Shah, "A survey of motion analysis from moving light displays," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pp. 214–221, Seattle, Wash, USA, June 1994.
- [12] R. G. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis and applications," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 781–796, 2000.
- [13] S. Niyogi and E. Adelson, "Analyzing gait with spatiotemporal surfaces," in *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 64–69, Austin, Tex, USA, November 1994.
- [14] S. Niyogi and E. Adelson, "Analyzing and recognizing walking figures in XYT," in *IEEE Computer Society Conference*



- on *Computer Vision and Pattern Recognition (CVPR '94)*, pp. 469–474, Seattle, Wash, USA, June 1994.
- [15] H. Murase and R. Sakai, "Moving object recognition in eigenspace representation: gait analysis and lip reading," *Pattern Recognition Letters*, vol. 17, no. 2, pp. 155–162, 1996.
  - [16] J. Little and J. Boyd, "Recognizing people by their gait: the shape of motion," *Videre*, vol. 1, no. 2, pp. 1–32, 1998.
  - [17] P. S. Huang, C. J. Harris, and M. S. Nixon, "Comparing different template features for recognizing people by their gait," in *British Machine Vision Conference*, pp. 639–648, Southampton, UK, 1998.
  - [18] Q. He and C. Debrunner, "Individual recognition from periodic activity using hidden Markov models," in *IEEE Workshop on Human Motion (HUMO '00)*, pp. 47–52, Austin, Tex, USA, December 2000.
  - [19] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Recognising human and animal movement by symmetry," in *Proc. IEEE Int. Conf. on Image Processing (ICIP '01)*, vol. 3, pp. 290–293, Thessaloniki, Greece, October 2001.
  - [20] D. Cunado, J. M. Nash, M. S. Nixon, and J. N. Carter, "Gait extraction and description by evidence-gathering," in *Audio- and Video-Based Biometric Person Authentication*, pp. 43–48, Washington, DC, USA, 1999.
  - [21] C. Yam, M. S. Nixon, and N. J. Carter, "Extended model-based automatic gait recognition of walking and running," in *Audio- and Video-Based Biometric Person Authentication*, pp. 278–283, Halmstad, Sweden, June 2001.
  - [22] A. F. Bobick and A. Y. Johnson, "Gait recognition using static activity-specific parameters," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. I-423–I-430, Kauai, Hawaii, USA, December 2001.
  - [23] P. C. Cattin, D. Zlatnik, and R. Borer, "Biometric system using human gait," in *Mechatronics and Machine Vision in Practice (M2VIP '01)*, Hong Kong, August 2001.
  - [24] J. Boyd, "Video phase-locked loops in gait recognition," in *IEEE International Conference on Computer Vision (ICCV '01)*, vol. 1, pp. 696–703, Vancouver, BC, Canada, July 2001.
  - [25] C. BenAbdelkader, R. Cutler, and L. Davis, "Motion-based recognition of people in eigengait space," in *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 254–259, Washington, DC, USA, May 2002.
  - [26] C. BenAbdelkader, R. G. Cutler, and L. S. Davis, "View-invariant estimation of height and stride for gait recognition," in *Post-ECCV Workshop on Biometric Authentication*, Copenhagen, Denmark, June 2002.
  - [27] A. Bissacco, A. Chiuseo, Y. Ma, and S. Soatto, "Recognition of human gaits," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 2, pp. 52–57, Kauai, Hawaii, USA, December 2001.
  - [28] A. Kale, A. N. Rajagopalan, N. Cuntoor, and V. Kruger, "Gait based recognition of humans using continuous HMMs," in *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 321–326, Washington, DC, USA, May 2002.
  - [29] L. Lee and W. E. L. Grimson, "Gait appearance for recognition," in *Post-ECCV Workshop on Biometric Authentication*, Copenhagen, Denmark, June 2002.
  - [30] R. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 351–356, Washington, DC, USA, May 2002.
  - [31] Y. Liu, R. Collins, and Y. Tsin, "Gait sequence analysis using frieze patterns," in *European Conference on Computer Vision (ECCV '02)*, pp. 657–671, Copenhagen, Denmark, May 2002.
  - [32] P. J. Philips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer, "Baseline results for the challenge problem of human ID using gait analysis," in *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 130–135, Washington, DC, USA, May 2002.
  - [33] Q. Cai and J. K. Aggarwal, "Human motion analysis: a review," in *Proc. IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects*, San Juan, Puerto Rico, June 1997.
  - [34] D. Gavrilu, "The visual analysis of human movement: A survey," *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82–98, 1999.
  - [35] S. Yasutomi and H. Mori, "A method for discriminating pedestrians based on rhythm," in *IEEE/RSG Int. Conf. on Intelligent Robots and Systems*, Munich, Germany, 1994.
  - [36] Y. Song, X. Feng, and P. Perona, "Towards detection of human motion," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '00)*, pp. 1810–1817, Hilton Head, SC, USA, 2000.
  - [37] L. W. Campbell and A. Bobick, "Recognition of human body motion using phase space constraints," in *International Conference on Computer Vision*, pp. 624–630, Cambridge, Mass, USA, 1995.
  - [38] J. W. Davis, "Appearance-based motion recognition of human actions," M.S. thesis, Media Arts and Sciences, MIT, September 1996.
  - [39] D. Meyer, J. Pösl, and H. Niemann, "Gait classification with HMMs for trajectories of body parts extracted by mixture densities," in *British Machine Vision Conference*, pp. 459–468, 1998.
  - [40] A. Kale, N. Cuntoor, and R. Chellappa, "A framework for activity-specific human recognition," in *International Conference on Acoustics Speech and Signal Processing*, Orlando, Fla, USA, May 2002.
  - [41] P. Tsai, M. Shah, K. Keiter, and T. Kasparis, "Cyclic motion detection for motion based recognition," *Pattern Recognition*, vol. 27, no. 12, pp. 1591–1603, 1994.
  - [42] J. W. Davis, "Visual categorization of children and adult walking styles," in *Audio- and Video-Based Biometric Person Authentication*, pp. 295–300, Halmstad, Sweden, June 2001.
  - [43] C. Yam, M. S. Nixon, and J. N. Carter, "Gait recognition by walking and running: a model-based approach," in *Asian Conference on Computer Vision*, pp. 1–6, Melbourne, Australia, 2002.
  - [44] C. BenAbdelkader, R. G. Cutler, and L. S. Davis, "Stride and cadence as a biometric in automatic person identification and verification," in *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 357–362, Washington, DC, USA, 2002.
  - [45] A. J. O'Toole, H. Abdi, K. Deffenbacher, and D. Valentin, "A perceptual learning theory of the information in faces," in *Cognitive and Computational Aspects of Face Recognition*, T. Valentin, Ed., chapter 8, pp. 159–182, Routledge, London, England, 1995.
  - [46] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face recognition: a literature survey," Tech. Rep. CAR-TR-948, UMD CFAR, 2000.
  - [47] R. G. Cutler, *On the detection, analysis, and applications of oscillatory motions in video sequences*, Ph.D. thesis, University of Maryland, College Park, Md, USA, 2000.
  - [48] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, New York Academic Press, New York, NY, USA, 1990.
  - [49] R. Duda, P. Hart, and D. Stork, *Pattern classification*, John Wiley & Sons, 2001.
  - [50] A. Elgammal, D. Harwood, and L. S. Davis, "Non-parametric model for background subtraction," in *International Conf. on Computer Vision (ICCV '99)*, Kerkira, Greece, September 1999.

- [51] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4s: A real-time system for detecting and tracking people in 21/2 d," in *European Conference on Computer Vision Computer Vision*, Freiburg, Germany, June 1998.
- [52] C. BenAbdelkader, R. G. Cutler, H. Nanda, and L. S. Davis, "Eigengait: Motion-based recognition of people using image self-similarity," in *Audio- and Video-Based Biometric Person Authentication*, Halmstad, Sweden, June 2001.
- [53] R. Polana and R. Nelson, "Detection and recognition of periodic, non-rigid motion," *International Journal of Computer Vision*, vol. 23, no. 3, pp. 261–282, 1997.
- [54] I. Haritaoglu, R. G. Cutler, D. Harwood, and L. S. Davis, "Backpack: Detection of people carrying objects using silhouettes," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 385–397, 2001.
- [55] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [56] I. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519–524, 1987.
- [57] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [58] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.
- [59] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [60] W. Zhao, R. Chellappa, and P. J. Phillips, "Subspace linear discriminant analysis for face recognition," Tech. Rep. CAR-TR-914, University of Maryland, 1999.
- [61] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [62] E. Borovikov, R. G. Cutler, T. Horprasert, and L. S. Davis, "Multi-perspective analysis of human actions," in *3rd International Workshop on Cooperative Distributed Vision*, Kyoto, Japan, November 1999.
- [63] R. Gross and J. Shi, "The CMU motion of body (MOBO) database," Tech. Rep. CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, June 2001.

**Ross G. Cutler** received his B.S. degree in mathematics, computer science, and physics in 1992; his M.S. degree in computer science in 1996; and his Ph.D. degree in computer science in 2000, in the area of computer vision from the University of Maryland, College Park. His research interests include multiview imaging, motion-based recognition, motion segmentation, HCI, video indexing, multimedia databases, gesture recognition, gait recognition, augmented reality, and real-time systems. He is currently a Researcher at Microsoft Corporation, working in the area of collaboration and multimedia systems. He has previously been employed at the University of Maryland, Johns Hopkins University, and has consulted for Emory University, University of Pennsylvania, Sony, and Sciencetech Inc.



**Larry S. Davis** received his B.A. degree from Colgate University in 1970 and his M.S. and Ph.D. degrees in computer science from the University of Maryland in 1974 and 1976, respectively. From 1977 to 1981, he was an Assistant Professor in the Department of Computer Science at the University of Texas, Austin. He returned to the University of Maryland as an Associate Professor in 1981. From 1985 to 1994, he was Director of the University of Maryland Institute for Advanced Computer Studies. He is currently a Professor in the Institute and the Computer Science Department, as well as Chair of the Computer Science Department. He was named a Fellow of the IEEE in 1997. Prof. Davis is known for his research in computer vision and high-performance computing. He has published over 75 papers in journals and has supervised over 15 Ph.D. students. He is an Associate Editor of the *International Journal of Computer Vision* and an Area Editor for *Computer Models for Image Processor: Image Understanding*. He has served as Program or General Chair for most of the field's major conferences and workshops, including the 5th International Conference on Computer Vision, the field's leading international conference.



**Chiraz BenAbdelkader** has received her B.S. and M.S. degrees in computer engineering from the Pennsylvania State University in 1994 and 1997, respectively, and a Ph.D. degree in computer science from the University of Maryland in 2002. Her Ph.D. dissertation was on automatic gait recognition from video sequences. She worked for a year from August 2002 to August 2003 as a Senior Research Scientist with Identix Corporation, a worldwide leader in biometrics technology specializing in face recognition and fingerprints. Currently, she is an Assistant Professor with the Department of Computer Science, American University of Beirut, Lebanon. Her research interests include human motion analysis in video, 3D face recognition, biometrics, and statistical pattern recognition.

