Human Gait Recognition with 3D Wavelets and Kernel based Subspace Projections

Nasir Rajpoot[1] Khalid Masood[2]

[1] Department of Computer Science, University of Warwick, UK
[2] COMSATS Institute of Information Technology, Pakistan

Abstract

Gait recognition can be regarded as a problem of uniquely representing spatiotemporal surfaces associated with a person's walking pattern in an efficient manner. In this paper, we describe the approach of using projections of such surfaces onto subspace spanned by appropriate axes using a single framework. Two new algorithms for gait recognition are presented which use projection on subspace of kernel induced higher dimensional spaces using PCA and Fisher's LDA. Wavelet transform in 3D is used to reduce the complexity of the problem. The proposed methods have been applied to datasets containing subjects walking at three different viewing angles in outdoor environment. The results show high accuracy results for recognizing subjects from their gait, even with a thumbnail size ($16 \times 16 \times 12$) of the gait patterns.

1 Introduction

Gait is the walking pattern of an individual. It has been shown [8, 2] that the human gait can be used as a biometric in passive surveillance applications. One of the reasons for it to be used as a biometric is that the human gait consists of synchronized integrated movements of hundreds of muscles and joints in an almost unique way for an individual [2]. Gait recognition is the task of identifying an individual from analysing their gait video. First, we present a brief review of the relevant work and explain our approach.

1.1 Relevant Work

A variety of methods for gait recognition can be found in the literature. Niyogi and Adelson [11] observed that the spatiotemporal behaviour of the human gait can be approximated by a smooth periodic surface. Indeed, a sequence containing a walking person can be processed (often simply by removing the background) to extract such periodic surfaces in 3D, assuming the background is static. In [11], such periodic surfaces are expressed as a combination of a standard parameterized surface, the so-called *canonical walk*, and a deviation surface which is more directly related to an individual's walk. Ben Abdelkader et al. [3] compute self-similarity plots (SSPs) from the spatiotemporal volume of a walking person by alignment and scaling of blobs containing silhouettes in each frame. Normalized SSPs are computed and standard classification algorithms are used in order to identify a person. Lee and Grimson [8] divide silhouettes of a walking person into seven regions and fit one ellipse to each region. Four parameters for each of the ellipses and one additional parameter for height of the silhouette make up their feature vector. Spatiotemporal analysis based on the Fourier transform is proposed by Ohara et al. [12], where Fourier transform is applied to the extracted gait volume and characteristic frequency of a walking person is determined. Three-dimensional (3D) kinematic models [4] have also been used to achieve good performance for individual recognition by their gait.

Perhaps of a more direct relevance to our work in this paper are subspace projection methods proposed in the literature. For instance, Mursae and Sakai [10] used PCA on the silhouette sequences to represent a person's gait by coordinates of projections on the subspace spanned by a few principal components. Huang et al. [6] represented gait in canonical space on temporal templates, recorded from optical flow changes between two consecutive spatial templates. Matching of the templates projected from high dimensional image space to low dimensional canonical space was shown to produce excellent results, although only one viewing angle is considered for training and testing. In a slightly different approach, Wang et al. [14] proposed to use normalized distance siganls, made up of distance values between the silhouette centroid and the boundary pixels, to approximate temporal patterns of gait. Principal component analysis (PCA) is applied to reduce the dimensionality of the distance signal sequences and a gait signature consisting of projections on the lower dimensional eigenspace is generated.

1.2 Our Approach

We perceive the problem of gait recognition as that of the representation of approximately periodic spatiotemporal surfaces in an efficient manner, preferably using only a few parameters. It is to be noted that the spatiotemporal surfaces associated with a person's gait normally consist of intertwined surfaces due to the motion of legs, torso, and the arms. The representation of such complex surfaces using measures such as gait frequency and phase is not sufficient to characterise a person's gait. We represent these surfaces as points in some high dimensional space $\Re^n$, where n is the size of the spatiotemporal volume containing the gait pattern. Based on the fact that the surfaces associated with human gait belong to a particular class of surfaces, our hypothesis is that these surfaces can be efficiently represented with a few parameters using projections onto appropriate axes spanning a subspace of $\Re^n$.

If a gait recognition system employing such a representation is to work, it will be required that gait surfaces of the same individual are mapped to points close by in the subspace, regardless of the viewpoint, illumination variations, person's mood, and speed of walking. Our work in this paper is motivated by the success of kernel based subspace projection methods for face recognition [15, 9]. Kernel based methods have also been succesful in other applications, see [13] for a detailed list. The basic idea behind support vector machines (SVMs) and other kernel based methods is that mapping the data from input space to a higher dimensional space often enables linear separation of data points belonging to different classes, which may otherwise not be linearly separable. We show that canonical space representation of the spatiotemporal gait patterns, when mapped onto a kernel induced higher dimensional space $\Re^f$ where $f > n$, is view-invariant and produces high recognition accuracy. Wavelet transform in 3D is used to obtain smaller resolution approximation of the gait patterns in order to reduce both storage and computational complexity.

In the next section, the training and recognition framework used in our work is presented. Building blocks of the framework are described in some detail. Experimental results are presented and discussed in Section 3, and conclusions and future directions are summarized in the final section.

2 The Training/Recognition Framework

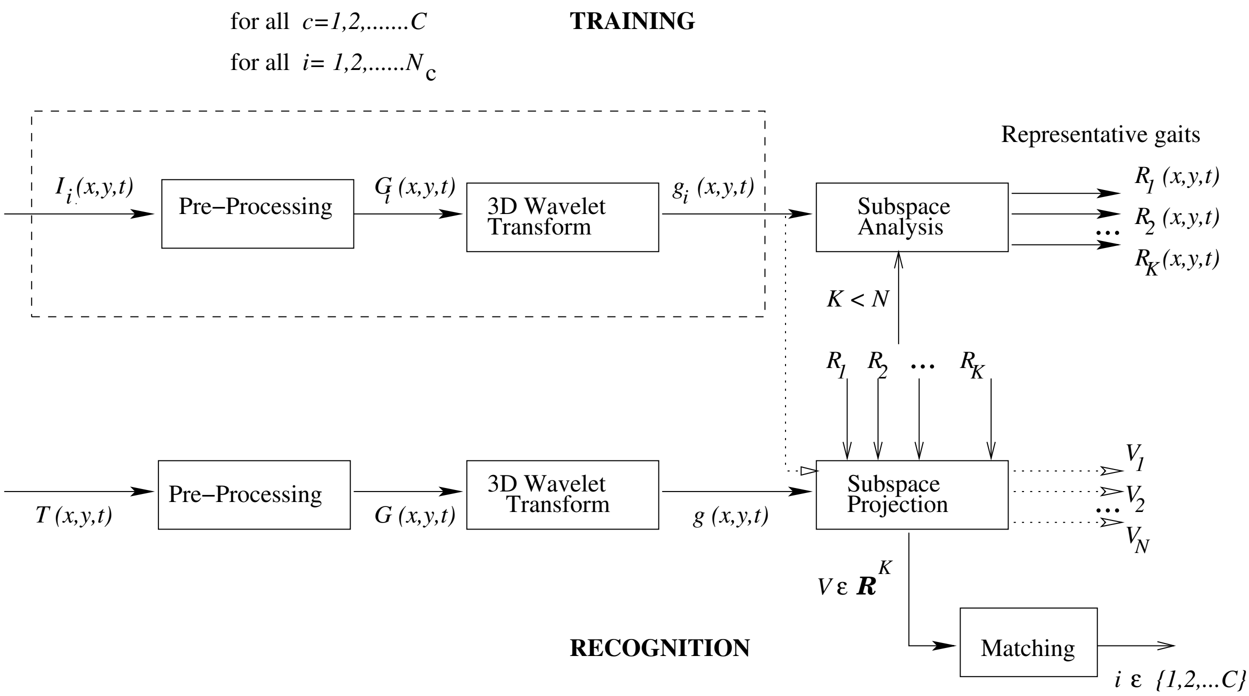

Figure 1: Block diagram of the proposed framework

A block diagram of the training/recognition framework employed in this work is shown in Figure 1. Let C and N_c respectively denote total number of subjects and number of training sequences for a particular subject c. Further, let N and K respectively denote the total number of sequences used for training and the number of *representative* gaits (i.e., the axes of subspace used for projection, principal vectors in case of PCA and generalized eigenvectors in case of LDA). At the **training** stage, representative gaits $R_i(x,y,t)$ for $i = 1,\ldots,K$ are computed using one of the subspace analysis methods (Section 2.3). Coordinates of the gait surfaces $G_i(x,y,t)$, $\forall i = 1,\ldots,N$, with respect to $R_i(x,y,t)$ are calculated and recorded as gait signatures. The **recognition** system takes as input a test sequence, computes the projection of its associated gait pattern onto the representative axes, and finds the closest match in terms of the normalized Euclidean distance.

2.1 Pre-Processing

The main purpose of the pre-processing stage is to extract spatiotemporal surfaces associated with gait pattern from the given sequence of a walking person. This is achieved by applying the following three modules:

i) *Background modelling*: A background image containing the scene information and without the walking figure can be reliably generated by taking the median of the image sequence. Assuming that the camera is stationary and only the object is moving, the background can be estimated as follows,

$$B(x,y) = \text{median}\{I(x,y,t)\} \tag{1}$$

where $1 \leq t \leq 60$ for each of the sequences.

ii) *Background subtraction*: In our experiments, we found that simple subtraction of the background through differencing between the background and the original frame, reversing the order, and adding the two subtracting images gives an equally good estimate of the foreground as compared to a relatively complex extraction function proposed in [7]. We chose to use the differencing approach due to its relatively low complexity.

iii) *Image enhancement*: Morphological erosion is used to filter spurious pixels and dilation is then used to fill any remaining holes. Histogram equalization and some contrast adjustment are performed before tracking each foreground region from frame to frame and placing a bounding box over the silhouette, with the size of the bounding box being the maximum of all the silhouette sizes. It is worth noting that unlike most other approaches to gait recognition, our silhouettes are not binary.

Figure 2: Results of pre-processing

Results of pre-processing for different views (the red rectangle is the bounding box): (a)–(c) fronto-parallel, (d)–(f) subject walking at 45° to the camera, (g)–(i) subject walking at 90° to the camera. Two-dimensional views of spatiotemporal surfaces $G_i(x,y,t)$ can be seen in (c), (f), and (i) for all three views.

2.2 3D Wavelet Decomposition

The pre-processed gait sequence $G(x,y,t)$ is decomposed into frequency subbands using a 3D wavelet transform. The lowest frequency subband gives a coarse approximation of the gait sequence, while the remaining subbands contain its contents corresponding to medium-to-high frequency phenomena in all three directions x, y, and t. Our experiments with subspace analysis on a number of combinations of different subbands of the 3D wavelet decomposition of $G(x,y,t)$ revealed that it was sufficient to perform subspace analysis on the lowest frequency subband. These findings are in agreement with an earlier study [5] which concluded that changes in angles or directions affect only the low-frequency spectrum and that only a change in subject will affect all frequency components. Hence we consider only a coarse approximation $g(x,y,t)$ of the gait sequence using the lowest frequency subband. Doing so not only substantially reduces the computational complexity of subsequent operations, it also allows us to accommodate the intermediate calculations (such as scatter matrices, see Section 2.3) in the memory.

2.3 Subspace Analysis/Projection

Subspace analysis can be regarded as the most important of all the components of training part of the framework in Figure 2. We employed four different methods for finding the representative gait patterns: principal component analysis (PCA), Fisher's linear discriminant analysis (LDA) [1], kernel PCA [13], and kernel LDA [9]. For the sake of completeness, here we provide an overview of all the methods in the context of the representation of spatiotemporal surfaces associated with an individual's gait.

2.3.1 PCA

PCA is a classical statistical method commonly used for dimensionality reduction. Given N gait sequences rearranged as n-dimensional points, PCA allows us to project the data points onto first K directions ($K < N$) while capturing the maximum variance of the data. The first K directions, which can be thought of as the first K major axes of an N-dimensional ellipsoid, are the eigenvectors corresponding to the largest K eigenvalues of the following covariance matrix:

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{g}_i - \overline{\mathbf{g}})(\mathbf{g}_i - \overline{\mathbf{g}})^T \tag{2}$$

where $\overline{\mathbf{g}}$ denotes the average of all N gait vectors.

2.3.2 LDA [1]

One of the problems with subspace projections using PCA is that it does not explicitly take into consideration any variation there may be between a class of gait patterns of the same subject. The LDA method, as described by Belhumeur et al. [1], uses both PCA and LDA to produce a subspace projection matrix, minimizing within-class variation and maximizing between-class variation. Two scatter matrices, within-class scatter $\mathbf{S}_w$ and between-class scatter $\mathbf{S}_b$, are computed for the training data set as follows.

$$\mathbf{S}_b = \sum_{c=1}^{C} N_c(\overline{\mathbf{g}}_c - \overline{\mathbf{g}})(\overline{\mathbf{g}}_c - \overline{\mathbf{g}})^T \tag{3}$$

$$\mathbf{S}_w = \sum_{c=1}^{C}\sum_{i=1}^{N_c}(\mathbf{g}_i - \overline{\mathbf{g}}_c)(\mathbf{g}_i - \overline{\mathbf{g}}_c)^T \tag{4}$$

where $\overline{\mathbf{g}}$ is as defined previously, and $\overline{\mathbf{g}}_c = \frac{1}{N_c}\sum_{i=1}^{N_c}\mathbf{g}_i$ is the average of each individual's gait patterns. If $\mathbf{S}_w$ is non-singular, the optimal matrix containing eigenvectors $\mathbf{W}_{opt}$ is chosen as the matrix with orthonormal columns which maximizes the following ratio,

$$\mathbf{W}_{opt} = \text{argmax}_{\mathbf{W}} \frac{|\mathbf{W}^T\mathbf{S}_b\mathbf{W}|}{|\mathbf{W}^T\mathbf{S}_w\mathbf{W}|} \tag{5}$$

where $\mathbf{W}_{opt} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{C-1}]$ is a matrix whose column vectors are generalized eigenvectors (forming the representative gaits for subspace projection during the recognition stage in Figure 2) of $\mathbf{S}_b$ and $\mathbf{S}_w$ corresponding to the K largest generalized eigenvalues $\{\lambda_i \mid i = 1, 2, \cdots, K\}$ satisfying the equation $\mathbf{S}_b\mathbf{w}_i = \lambda_i\mathbf{S}_w\mathbf{w}_i$.

2.3.3 Kernel PCA [13]

The basic assumption behind standard PCA is that the probability distribution in the n-dimensional gait space is a multidimensional Gaussian, making it possible to reduce the dimensionality of the problem by projecting the data onto first few principal axes capturing a large amount of the variation among the data. In case of non-Gaussian distributions, one way to deal with the problem is to make the data in the gait space more Gaussian-like (and hence more easily separable) by nonlinearly mapping it to a higher dimensional space $\Re^f$ using an appropriate map $\Phi(.) : \Re^n \to \Re^f$, with $f > n$.

The so-called *kernel trick* allows us to reduce the complexity of the mapping by using a kernel which when applied to two data points is equivalent to the dot product of the data points mapped to $\Re^f$. PCA on $\Re^f$ can be applied as follows. Let $\mathbf{C}^\Phi$ denote the covariance matrix of data points from the training set mapped onto $\Re^f$. In order to compute the eigenvectors in $\Re^f$, the following eigenvalue problem needs to be solved

$$\lambda\mathbf{w}^\Phi = \mathbf{C}^\Phi\mathbf{w}^\Phi \tag{6}$$

where $\mathbf{w}^\Phi$ denotes eigenvector in $\Re^f$ and for all non-zero eigenvalues, it lies in the span of $\Phi(\mathbf{g}_1), \Phi(\mathbf{g}_2), \cdots, \Phi(\mathbf{g}_N)$. In other words,

$$\mathbf{w}^\Phi = \sum_{i=1}^{N} \alpha_i \Phi(\mathbf{g}_i) \tag{7}$$

If we denote by $\mathbf{K}$ the kernel matrix whose elements are given by

$$K_{ij} = k(\mathbf{g}_i, \mathbf{g}_j) = \Phi(\mathbf{g}_i) \cdot \Phi(\mathbf{g}_j) \;, \tag{8}$$

it can be shown that the eigenvalue problem of (6) is equivalent to the following eigenvalue problem [13],

$$N\lambda\alpha = \mathbf{K}\alpha. \tag{9}$$

Solving the above equation yields eigenvalues λ and corresponding eigenvectors α consisting of N elements. Importantly, it does not requires explicit computation of the mapping Φ. A knowledge of the kernel matrix $\mathbf{K}$ suffices.

Given a test gait pattern $\mathbf{g}$, its projection onto an eigenvector $\mathbf{w}^\Phi$ of $\Re^f$ can be calculated using the following equation:

$$\mathbf{w}^\Phi \cdot \Phi(\mathbf{g}) = \sum_{i=1}^{N} \alpha_i \; (\Phi(\mathbf{g}_i) \cdot \Phi(\mathbf{g})) = \sum_{i=1}^{N} \alpha_i \; k(\mathbf{g}_i, \mathbf{g}) \tag{10}$$

Extracting the first K eigenvectors $\mathbf{w}^\Phi$ gives nonlinear principal components using the kernel function without explicitly mapping onto $\Re^f$. In the gait space, first K eigenvectors of (9) give representative gaits which can be used for projections as above.

2.3.4 Kernel LDA [9]

Similar to kernel PCA, kernel LDA in its standard form is the application of LDA to the high-dimensional space $\Re^f$. In other words, it requires the solution to the following generalized eigenvalue problem,

$$\lambda \mathbf{S}_w^\Phi \mathbf{w}^\Phi = \mathbf{S}_B^\Phi \mathbf{w}^\Phi \tag{11}$$

which can be obtained by

$$\mathbf{W}_{opt}^\Phi = \text{argmax}_{\mathbf{W}^\Phi} \frac{|(W^\Phi)^T S_b^\Phi W^\Phi|}{|(W^\Phi)^T S_w^\Phi W^\Phi|} = [\mathbf{w}_1^\Phi, \mathbf{w}_2^\Phi, \cdots, \mathbf{w}_m^\Phi]. \tag{12}$$

Once again, any solution $\mathbf{w}^\Phi \in \Re^f$ should lie in the span of all training samples in $\Re^f$. Mathmetically,

$$\mathbf{w}^\Phi = \sum_{c=1}^{C} \sum_{i=1}^{N_c} \alpha_{ci} \Phi(\mathbf{g}_i) \tag{13}$$

It can be shown that the solution to above equation can be found by solving the following problem [15]:

$$\lambda \mathbf{K}\mathbf{K}\alpha = \mathbf{K}\mathbf{Z}\mathbf{K}\alpha \tag{14}$$

where $\mathbf{Z} = (\mathbf{Z}_c)_{c=1,\cdots,C}$ and $(\mathbf{Z}_c)$ is a $N_c \times N_c$ matrix with all its terms equal to $1/N_c$. Projections onto eigenvectors $\mathbf{w}^\Phi$ can be computed in the same way as in kernel PCA.

3 Experimental Results

In this section, we compare the performance of all the four subspace projection techniques on pre-processed training and test sequences containing a walking person. We selected NLPR dataset because of its complexity in viewing directions and use of outdoor environment. Our experiments were conducted on 15 subjects walking at three different angles relative to the camera, as shown in Figure 2.1, recorded at two different times on two different days. The data contains subjects walking at different speeds in two opposite directions for each of the angles. Half of the dataset, that is six sequences per subject (two per angle), is taken as the training set and the other half is used for the test purposes. The pre-processing resulted in gait sequences $\mathbf{G}_i(x,y,t)$, each of dimensions $64 \times 64 \times 48$. Coarse approximation for each of the gait sequences was obtained by extracting the lowest frequency subband after performing a two-level wavelet transforum using Daubechies-4 filters, giving $16 \times 16 \times 12$ sequence $g_i(x,y,t)$ to be used for subspace analysis (projection) during training (testing). Kernel PCA uses a polynomial kernel of degree 2, which gave better accuracy as compared to the Gaussian kernel in our experiments. Kernel LDA is performed with tuning of polynomial kernel of degree 3, which produced the best results.

In order to have a quantitative evaluation of the performance, normalized coordinates of each test subject sequence is compared with the sequences in the training set using the normalized Euclidean distance. The recognition accuracy for each technique is defined as the ratio of the number of sequences correctly identified to the total number of sequences in the tests. A graph of recognition accuracy against the number of subspace components K is shown in Figure 3. The value of K is varied from 4 to 30 for PCA and kernel PCA, whereas the maximum value of K is one less than the number of subjects in the training set for both LDA and kernel LDA. The best results for all four algorithms are shown in Table 1.

Based on these results, we make the following observations. Kernel LDA is the overall winner in terms of accuracy, although it is computationally intensive, has high memory requirements, and requires careful tuning. PCA is relatively faster in computation but least accurate. Kernel PCA performance is closely matched by PCA, similar to the results in [15], perhaps due to the inclusion of difficult viewing angles. Nevertheless, the fact that both the training and test data contained three different viewing angles shows that the results are promising. Extensive experimentation using different datasets is the next logical step of our research.

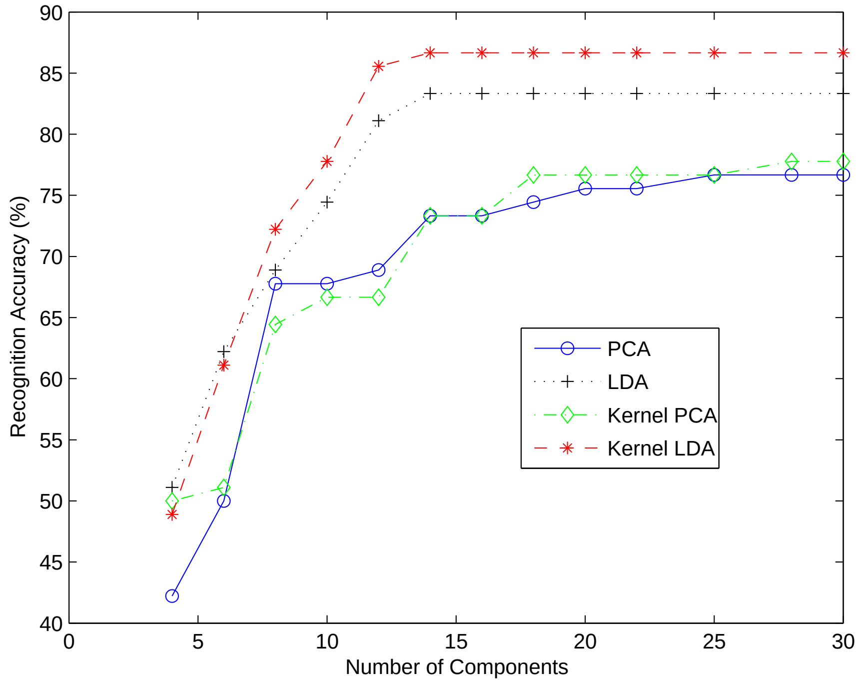

Figure 3: Comparative Results of Four Subspace Projection Methods

Method	K	Kernel	Accuracy (%)
PCA	20	-	75.5
Kernel PCA	20	polynomial (order 2)	76.6
LDA	14	-	83.3
Kernel LDA	14	polynomial (order 3)	86.6

Table 1: Best recognition results using the four methods

4 Conclusions

In this paper, we presented a comparison of four subspace projection methods for gait recognition. After pre-processing, the gait sequence was approximated at a coarse resolution using 3D wavelet transform thus reducing the storage and computational complexity of the subsequent subspace analysis. Two of the methods were based on the subspace analysis of the kernel induced higher dimensional space, which have not previously been reported in the literature to the best of our knowledge. Kernel LDA produces the best results using a polynomial kernel, while LDA offers a relatively less expensive alternative with competetive performance. Our results show that view-invariant representation of the spatiotemporal surfaces associated with a person's gait may be achievable using kernel based analysis methods. It is our belief that subspace projection can also be used to recognize certain human activities. More research, however, is needed to improve robustness of these methods.

Acknowledgements

The authors would like to thank the NLPR administrators for providing access to their gait database used in this work.

References

[1] P. Belhumeur, J. Hesponha, and D. Kriegman. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18, 1996.

[2] C. BenAbdelkader, R. Culter, and L. Davis. Gait recognition using image self-similarity. *EURASIP Journal on Applied Signal Processing*, 4(4):1–14, 2004.

[3] C. BenAbdelkader, R. Culter, H. Nanda, and L.Davis. EigenGait: Motion based recognition of people using image self-similarity. In *Proc. Intl. Conf. Audio- and Video-Based Biometric Person Authentication*, 2001.

[4] B. Bhanu and J. Han. Individual recognition by kinematic-based gait analysis. In *Proc. IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, 2002.

[5] G.C. Feng, P.C. Yuen, and D.Q. Dai. Human face recognition using PCA on wavelet subband. *Journal of Electronic Imaging*, 9(2):226–233, 2000.

[6] P.S. Huang, C.J. Harris, and M.S. Nixon. Human gait recognition in canonical space using temporal templates. *IEE Proc. Vision and Image Signal Processing*, 1999.

[7] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S.Nakagawa. Automated detection of human for visual surveillance systems. In *Proc. IEEE Intl. Conf. Pattern Recognition*, 1996.

[8] L. Lee and W. Grimson. Gait analysis for recognition and clasification. In *Proc. IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, 2002.

[9] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithm. *IEEE Trans. Neural Networks*, 2002.

[10] H. Mursae and R. Sakai. Moving object recognition in eigen-space representation: Gait analysis and lip reading. *Pattern Recognition Letters*, 12, 1996.

[11] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in XYT. In *Proc. IEEE Computer Vision and Pattern Recognition*, 1994.

[12] Y. Ohara, R. Sagawa, T. Echigo, and Y. Yogi. Gait volume: Spatio-temporal analysis of walking. In *Proc. MFI*, 2003.

[13] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[14] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2003.

[15] M.-H. Yang. Kernel eigenfaces vs kernel fisherfaces: Face recognition using kernel methods. In *Proc. IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, 2002.