Comparing Different Template Features for Recognizing People by Their Gait

Ping S. Huang, Chris. J. Harris and Mark S. Nixon
Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, UK
[psh95r|cjh|msn]@ecs.soton.ac.uk

Abstract

To recognize people by their gait from a sequence of images, we have proposed a statistical approach which combined *eigenspace transformation* (EST) with *canonical space transformation* (CST) for feature transformation of spatial templates. This approach is used to reduce data dimensionality and to optimize the class separability of different gait sequences simultaneously. Good recognition rates have been achieved. Here, we incorporate temporal information from optical flows into three kinds of temporal templates and use them as features for gait recognition in addition to the spatial templates. The recognition performance for four kinds of template features has been evaluated in this paper. Experimental results show that spatial templates, horizontal-flow templates and the combined horizontal-flow and vertical-flow templates are better than vertical-flow templates for gait recognition.

1 Introduction

Biometrics such as automatic face and voice recognition continue to be subject to increasing interest. Gait is a new biometric aimed to recognize subjects by the way they walk [6, 10, 8, 3]. Recently, Niyogi and Adelson [11] distinguished different walkers by extracting their spatio-temporal gait patterns obtained from the curve-fitting "snakes". Cunado *et al.* [3] developed a technique which considers legs as an interlinked penduli and use phase-weighted Fourier magnitude spectra as the feature to recognize different people. Little and Boyd [8] use frequency and phase features from optical flow information to recognizeeigenface different people by their gait. However, these feature-based methods, which use boundaries, lines, edges or optical flow are dependent on the reliability of the feature extraction process. Using the human shapes and their temporal changes during walking, Murase and Sakai [10] proposed a template-matching method which uses the parametric eigenspace representation, as applied in face recognition [12], to recognize different human gait. For recognizing people by their gait, this appears more potent compared with other approaches. Based on Principal Component Analysis (PCA), eigenspace transformation (EST) has actually been demonstrated to be a potent metric in automatic face recognition and gait analysis, but without using data analysis to increase classification capability.

BMVC 1998 doi:10.5244/C.12.64

Previously, we proposed a statistical approach [6] which combined *eigenspace transformation* (EST) with *canonical space transformation* (CST) based on Canonical Analysis for feature extraction of spatial templates to recognize humans by their gait. This approach can be used to reduce data dimensionality and to optimize the class separability of different gait sequences simultaneously. By using spatial templates of human silhouettes as features, each image template is projected from high-dimensional image space to a single point in low-dimensional canonical space. A walking sequence becomes a trajectory in this new space and the recognition of human gait becomes much simpler and more accurate.

In this paper we incorporate temporal information from optical-flow changes between two consecutive spatial templates into three kinds of temporal templates and use them as additional features for gait recognition. Firstly, spatial and temporal templates are extracted from each sequence. Secondly, training template sequences are projected into individual canonical space by EST and CST after training. Thirdly, recognition of test sequences is achieved in canonical space after projection. Finally, the recognition performance of the four kinds of template features is shown in experimental results.

2 Feature Template Extraction

Intuitively, recognizing humans by gait depends on how the silhouette changes for individual subjects. According to this hypothesis, here we use spatial templates in [6] and three kinds of temporal templates to recognize humans by gait. Before training and recognition, each gait sequence is separately converted into four template sequences in which spatial templates and three kinds of temporal templates are extracted from each original sequence.

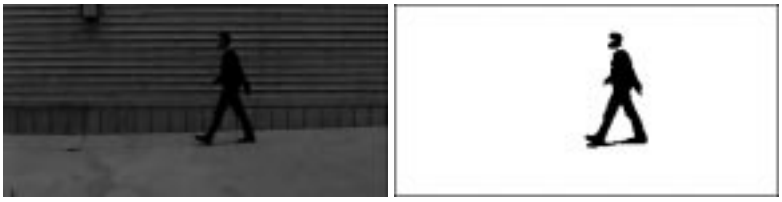
2.1 Spatial Templates

For the extraction of spatial templates, we choose the preliminary process from Murase's approach [10] in which the silhouette is fitted in a 64×64 image template by normalizing its position and size with constant aspect ratio. Naturally, to isolate the human silhouette, we can simply subtract the background from each image. However, the difference image thus obtained is not binarized. To simplify the representation, a binary image is obtained by region growing [4]. Figure 1(a) shows an image from a gait sequence and Figure 1(b) is a binary(thresholded) version.

The centroid and silhouette window of each template in the original image can be obtained simultaneously and are used later for the extraction of temporal templates. Sample spatial templates are illustrated in Figure 2 from a gait sequence.

2.2 Temporal Templates

For the extraction of temporal templates, Little and Boyd's [8] technique which based on the algorithm of Bulthoff *et al.* [2] is used to generate optical flow fields between two consecutive images. Instead of isolating the moving figure manually, as in [8], we use the information of centroid and silhouette window from the extraction of spatial templates to



(a) A walking person (b) thresholded version

Figure 1: Sample images of a walking person

(a) (b) (c) (d) (e) (f)

Figure 2: Sample spatial templates of a walking person

extract each temporal template which contains the flow within a moving window. Figures 3(a) and 3(b) show two consecutive images from a gait sequence.

Unlike other methods, Little and Boyd [8] used dense optical flow fields, generated by minimizing the sum of absolute differences between image patches [2]. However, this algorithm is sensitive to brightness change caused by reflections, shadows, and changes of illumination. Therefore, the images are firstly processed by computing the logarithm of brightness and converting the multiplicative effect of illumination change into an additive one. Secondly, each processed image is filtered by a bandpass filter (Laplacian of Gaussian) to remove the additive effects.

Basically, the algorithm [2] searches for the displacement of each pixel among a limited set of discrete displacements by minimizing the sum of absolute differences between

(a) Current image (b) Next image

Figure 3: Two consecutive images

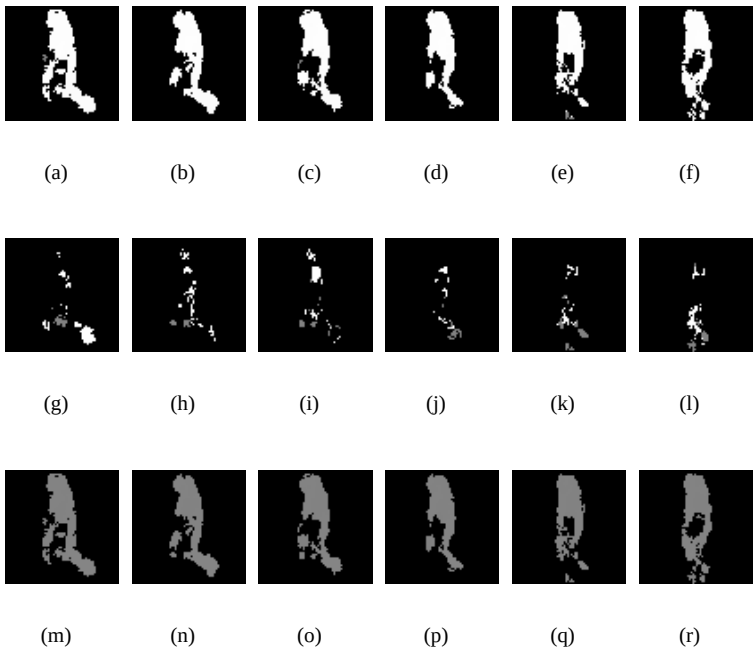


Figure 4: Sample temporal templates of a walking person

a patch in one image and the corresponding displaced patch in the other image. After a best matching patch in the second image is found for each patch in the first, the algorithm is run a second time and the roles of the the two images are switched. For a correct match, the results will likely agree. In order to remove invalid matches, Little and Boyd compare the results at each point in the first image with the result at the corresponding point in the second. The second point should match to the first: the sum of displacement vectors should be approximately zero. Only those matches that pass this validation test are retained. The results could be interpolated to provide sub-pixel displacements but only integral values are used here. In effect, the minimum displacement is 1.0 pixels per frame; points that are assigned non-zero displacements form a set of *moving points*.

Three kinds of temporal templates are adopted in this paper and they are u-flow templates which are horizontal components of flow, v-flow templates which are vertical components of flow and $|(u, v)|$-flow templates which are the magnitudes of (u, v). They are shown in Figures 4. Figure 4(a)-(f) are u-flow templates, Figure 4(g)-(l) are v-flow templates and Figure 4(m)-(r) are $|(u, v)|$-flow templates from a gait sequence, respectively. For display purposes, stationary pixels are represented by gray-value 0, positive components are offset by 128 and negative components by subtracting its absolute value from 255.

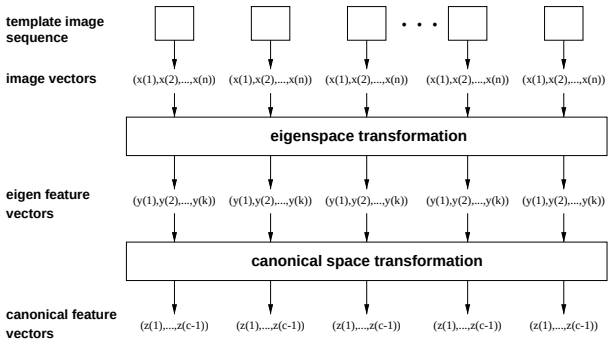


Figure 5: Projection of template images

3 Transformation and Training

Four kinds of basic feature templates from each training sequence are used during training, they are spatial templates, u-flow templates, v-flow templates and $|(u,v)|$-flow templates, respectively. They will be individually projected into four different canonical spaces by for further recognition after training stage. Figure 5 illustrates the projection steps that generate feature vectors by eigenspace transformation and canonical space transformation for different kinds of template images.

Adopted from previous work [6], we use the transformation which combined EST and CST for feature extraction. Template images in high-dimensional image space are converted to low-dimensional eigenspace using EST. Obtained vectors thus are further projected to a smaller canonical space using CST. Recognition is accomplished in canonical space. Patently, the reduced dimensionality results in concomitant decrease in computation cost.

Given c classes for training and each class represents a template sequence of a single person. $\mathbf{x}_{i,j}$ is the j-th template in class i, and N_i is the number of templates in i-th class, the total number of training templates is $N_T = N_1 + N_2 + \cdots + N_c$. This training set is represented by

$$\left[\mathbf{x}_{1,1}, \cdots, \mathbf{x}_{1,N_1}, \mathbf{x}_{2,1}, \cdots, \mathbf{x}_{c,N_c}\right] \tag{1}$$

where each sample $\mathbf{x}_{i,j}$ is an template image with n pixels. By subtracting the mean $\mathbf{m}_\mathbf{x}$ of full image set from each image, the image set can be described by a $n \times N_T$ matrix $\mathbf{X}$, with each image $\mathbf{x}_{i,j}$ forming one column of $\mathbf{X}$, that is

$$\mathbf{X} = [\mathbf{x}_{1,1} - \mathbf{m}_\mathbf{x}, \cdots, \mathbf{x}_{1,N_1} - \mathbf{m}_\mathbf{x}, \cdots, \mathbf{x}_{c,N_c} - \mathbf{m}_\mathbf{x}]. \tag{2}$$

3.1 Eigenspace Transformation

Let $\mathbf{R}$ be a $n \times n$ matrix and represented by

$$\mathbf{R} = \mathbf{X}\mathbf{X}^\mathbf{T}. \tag{3}$$

Based on *singular value decomposition* theory [9], eigenvalues and associated eigenvectors of $\mathbf{R}$ can be recovered from a much smaller matrix,

$$\tilde{\mathbf{R}} = \mathbf{X}^\mathbf{T}\mathbf{X}, \tag{4}$$

by the relationships

$$\begin{cases} \lambda_i = \tilde{\lambda}_i \\ \mathbf{e}_i = \tilde{\lambda}_i^{-\frac{1}{2}} \mathbf{X} \tilde{\mathbf{e}}_i \end{cases}, \tag{5}$$

where $i = 1, \ldots, K$, $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_K$ and $\tilde{\mathbf{e}}_i, \ldots, \tilde{\mathbf{e}}_K$ are eigenvalues and eigenvectors of $\tilde{\mathbf{R}}$. Suppose obtained eigenvalues and associated eigenvectors of $\mathbf{R}$ are $\lambda_1, \ldots, \lambda_K$ and $\mathbf{e}_i, \ldots, \mathbf{e}_K$. According to the theory of PCA, each image can be approximated by taking only the $k \leq K$ largest eigenvalues $\mid \lambda_1 \mid \geq \mid \lambda_2 \mid \geq \cdots \geq \mid \lambda_k \mid$ and associated eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$. This partial set of k eigenvectors spans an eigenspace in which $\mathbf{y}_{i,j}$ are the points that are the projections of the original images $\mathbf{x}_{i,j}$ by the equation

$$\mathbf{y}_{i,j} = [\mathbf{e}_1, \cdots, \mathbf{e}_k]^{\mathbf{T}} \mathbf{x}_{i,j}, \tag{6}$$

where $i = 1, \ldots, c$ and $j = 1, \ldots, N_c$.

3.2 Canonical Space Transformation

Based on the theory of canonical analysis [5], CST is presented as follows. Suppose $\{\Phi_1, \Phi_2, \cdots, \Phi_c\}$ represents the classes of transformed vectors by eigenspace transformation and $\mathbf{y}_{i,j}$ is the j-th vector in class i. The mean vector of the entire set is given by

$$\mathbf{m}_{\mathbf{y}} = \frac{1}{N_T} \sum_{i=1}^{c} \sum_{j=1}^{N_i} \mathbf{y}_{i,j}, \tag{7}$$

and the mean vector of the i-th class is represented by

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{\mathbf{y}_{i,j} \in \Phi_i} \mathbf{y}_{i,j}. \tag{8}$$

Let $\mathbf{S_w}$ denote *within-class matrix* and $\mathbf{S_b}$ denote *between-class matrix*, then

$$\begin{aligned} \mathbf{S_w} &= \frac{1}{N_T} \sum_{i=1}^{c} \sum_{\mathbf{y}_{i,j} \in \Phi_i} (\mathbf{y}_{i,j} - \mathbf{m}_i)(\mathbf{y}_{i,j} - \mathbf{m}_i)^{\mathbf{T}} \\ \mathbf{S_b} &= \frac{1}{N_T} \sum_{i=1}^{c} N_i (\mathbf{m}_i - \mathbf{m}_{\mathbf{y}})(\mathbf{m}_i - \mathbf{m}_{\mathbf{y}})^{\mathbf{T}} \end{aligned}$$

The objective is to minimize $\mathbf{S_w}$ and maximize $\mathbf{S_b}$ simultaneously, that is to solve the *generalized eigenvalue equation*

$$\mathbf{S_b} \mathbf{w}_i^* = \lambda_i \mathbf{S_w} \mathbf{w}_i^*. \tag{9}$$

After equation (9) is solved, we will obtain $(c-1)$ nonzero eigenvalues and their corresponding eigenvectors $[\mathbf{v}_1, \ldots, \mathbf{v}_{c-1}]$ that create another orthogonal basis and span a $(c-1)$-dimensional canonical space. By using this basis, each point in eigenspace can be further projected to another point in this canonical space by

$$\mathbf{z}_{i,j} = [\mathbf{v}_1, \cdots, \mathbf{v}_{c-1}]^{\mathbf{T}} \mathbf{y}_{i,j}, \tag{10}$$

where $\mathbf{z}_{i,j}$ represents the new point and $[\mathbf{z}_{i,1}, \ldots, \mathbf{z}_{i,N_i}]$ is the new trajectory in canonical space. By merging equation (6) and equation (10), each image can be projected into one point in the new $(c-1)$-dimensional space by

$$\mathbf{z}_{i,j} = [\mathbf{v}_1, \ldots, \mathbf{v}_{c-1}]^{\mathbf{T}} [\mathbf{e}_1, \ldots, \mathbf{e}_k]^{\mathbf{T}} \mathbf{x}_{i,j}. \tag{11}$$

The *centroid* of each training sequence in canonical space is given by

$$\mathbf{C}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{z}_{i,j} \tag{12}$$

4 Recognition

Let a test gait sequence be $g(t)$, in which $t = 1, ..., T$. Before recognition, four different kinds of templates are extracted from this test sequence and projected into individual trained canonical space by equation (11), given four vector sequences after projection, $h_1(t)$, $h_2(t)$, $h_3(t)$ and $h_4(t)$, representing spatial, u-flow, v-flow and $|(u, v)|$-flow templates, respectively.

To recognize a human walking sequence from a trained database in each canonical space, the *accumulated distance to each centroid* is used. This will eliminate matching problems caused by velocity changes and phase shifts. The accumulated distance between test vector sequences, $h_k(t)$, in which $k = 1, ..., 4$ and six centroids, $\mathbf{C}_{i,k}$, in which $i = 1, ..., 6$ is

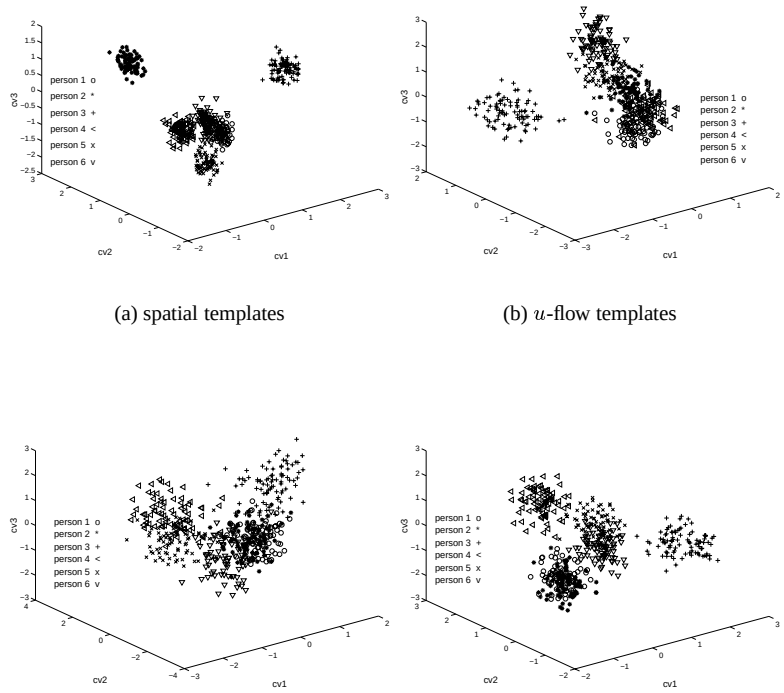
$$d_{i,k}^2 = \sum_{t=1}^{T} \|h_k(t) - \mathbf{C}_{i,k}\|^2, \tag{13}$$

where $\mathbf{C}_{i,k}$ is the centroid of class i in canonical space k. To match a test sequence $h_k(t)$ to a training sequence i in canonical space k can be accomplished by choosing the *minimum* $d_{i,k}^2$.

5 Experimental Results

The sample human gait data came from the Visual Computing Group, University of California, San Diego. There are 6 people and 7 sequences of each. One walking sequence is selected from each person as the training sequence and remaining 36 sequences served as test sequences. Results in Figures 6(a), 6(b), 6(c) and 6(d) show that six classes of training sequences using spatial templates, u-flow templates, v-flow templates and $|(u, v)|$-flow templates are greatly separated in each canonical space. For visualization purposes, we only show the first three of five dimensions. Linear re-scaling [1] has been applied to each vector to set the average of each data set to zero and to normalize the standard deviation to unity.

Figures 7(a) and 7(b) show relative accumulated distances of one training and one test sequences from subject 1. Here, the v-flow template has lower distance and hence poorest discriminatory ability. Conversely, the spatial template offers best discriminatory ability, associated with greatest distance. Figures 7(c) and 7(d) show relative accumulated distances of two misclassified sequences, one from subject 4 and one from subject 5. Here,



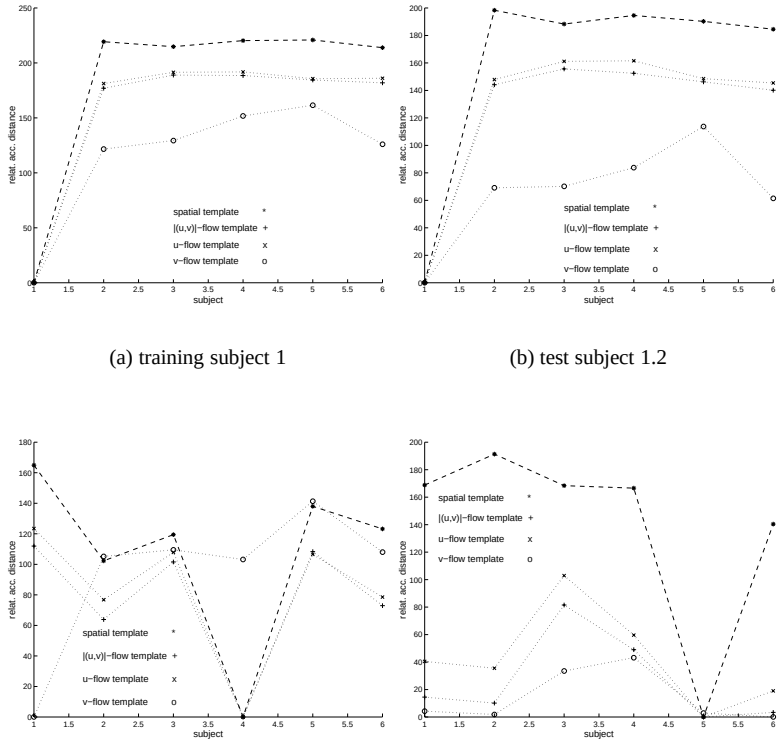
(a) spatial templates (b) u-flow templates

(c) v-flow templates (d) $|(u, v)|$-flow temporal templates

Figure 6: Distributions of 6 training sequences in four canonical spaces

the distance by the v-flow template leads to confusion in classification, in Figure 7(c), where the fifth sequence of subject 4 can be classified as subject 1 and, in Figure 7(d), subjects 1 and 2 appear close to the target subject 5. Conversely, the spatial template again offers best performance, again there is little difference between the $|(u, v)|$-flow template and the u-flow template and both perform better than the v-flow template measures.

The comparison of recognition performance using four different templates is shown in Table 1. Clearly, the feature vectors generated by the combination of EST and CST yield high recognition rates. Using template matching, the poor performance achieved by v-flow templates can be explained by the reduced information of optical flow from the extracted templates in Figure 4(g)-(l). Vertical movements of gait usually have smaller changes than horizontal movements, thus have less discriminatory power in distinguishing different gaits. Spatial templates, u-flow templates and $|(u, v)|$-flow templates make better performance in recognition. Although promising results have been shown here, further comparison of the three templates still needs a larger database to better assess their performance.



(a) training subject 1 (b) test subject 1.2

(c) test subject 4.5 (d) test subject 5.6

Figure 7: Relative accumulated distance of training and test sequences

6 Conclusions

In this paper we use our previous approach which combined EST with CST for feature extraction with new motion data. EST ans CST can be used to reduce data dimensionality and to optimize the class separability of different classes simultaneously, greatly improving the performance of eigenspace approach. Apart from the feature of spatial templates used in previous work, we propose three more features by temporal templates. These new features incorporate temporal information into each template. The analysis and comparison of recognition performance for each individual feature shows that the spatial templates, the horizontal u-flow templates and the magnitude $|(u, v)|$-flow templates are better than the vertical v-flow templates for gait recognition. We are currently working on extension of the test environment. In addition to testing on larger database, the extended features by different combinations of four kinds of templates will be also evaluated. For extended feature vectors, it has been also suggested in [7] that orthogonal feature sets should be chosen to reduce the variance of a final match measure. The combined feature of spatial and temporal templates incorporates spatial and temporal information into one single feature, its robustness is worthy of further investigation. Future work will also

	feature used	recognition rate
(1)	spatial templates	100%
(2)	u-flow templates	100%
(3)	v-flow templates	95.2%
(4)	$\|(u, v)\|$-flow templates	100%

Table 1: recognition using different template features

concentrate on looking for more precise and robust features, whilst aiming to develop the technique still further.

7 Acknowledgements

We would like to thank Dr. Jeffrey Boyd at the Visual Computing Laboratory, University of California, San Diego, USA for providing gait data and giving invaluable advice.

References

[1] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1996.

[2] H. Bulthoff, J. Little, and T. Poggio. A parallel algorithm for real-time computation of optical flow. *Nature*, 337:549–553, February 1989.

[3] D. Cunado, M.S. Nixon, and J.N. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *First Int. Conf., AVBPA'97*, pages 95–102, Crans-Montana, Switzerland, March 1997.

[4] M-P Dubuisson and A.K. Jain. Contour extraction of moving objects in complex outdoor scenes. *International Journal of Computer Vision*, 14(6):83–105, 1995.

[5] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2nd edition, 1990.

[6] P.S. Huang, C.J. Harris, and M.S. Nixon. Canonical space representation for recognizing humans by gait and face. In *Proc. of Southwest Symposium on Image Analysis and Interpretation*, pages 180–185, Tucson, Arizona, USA, April 1998. IEEE.

[7] X. Jia and M.S. Nixon. Extending the feature vector for automatic face recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(12):1167–1176, 1995.

[8] J. Little and J. Boyd. Recognizing people by their gait: the shape of motion. *MIT Press Journal - Videre*, 1997. Accepted for publication.

[9] H. Murakami and V. Kumar. Efficent calculation of primary images from a set of images. *IEEE Trans. on Pattern Anal. Machine Intell.*, 4(5):511–515, 1982.

[10] H. Murase and R. Sakai. Moving object recognition in eigenspace representation: gait analysis and lip reading. *Pattern Recognition Letters*, 17:155–162, 1996.

[11] S.A. Niyogi and E.H. Adelson. Analysis and recognizing walking figures in xyt. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 469–474, Seattle, WA, USA, June 1994.

[12] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.