Human gait recognition in canonical space using temporal templates

P.S. Huang, C.J. Harris and M.S. Nixon

Abstract: A system for automatic gait recognition without segmentation of particular body parts is described. Eigenspace transformation (EST) has already proved useful for several tasks including face recognition, gait analysis, etc; it is optimal in dimensionality reduction by maximising the total scatter of all classes but is not optimal for class separability. A statistical approach that combines EST with canonical space transformation (CST) is proposed for gait recognition using temporal templates from a gait sequence as features. This method can be used to reduce data dimensionality and to optimise the class separability of different gait sequences simultaneously. Incorporating temporal information from optical-flow changes between two consecutive spatial templates, each temporal template extracted from computation of optical flow is projected from a high-dimensional image space to a single point in a low-dimensional canonical space. Using template matching, recognition of human gait becomes much faster and simpler in this new space. As such, the combination of EST and CST is shown to be of considerable potential in an emerging new biometric.

1 Introduction

Biometrics are methods to automatically recognise a person by a physiological or behavioural characteristics. Examples of human traits currently used for biometric recognition include fingerprints, speech and face. Handwritten signatures are used for verification. Fingerprint and face recognition have already been used in commercial and law enforcement applications. Gait is a new biometric aimed at recognising subjects by the way they walk.

The earliest approach to recognising certain subjects by their gait was proposed by Cutting and Kozlowski [1] in the early 1970s. They used a retroreflective material wrapped around walker's joints. Viewers recognised their friends from the video-taped images. Using similar techniques, Johansson [2] shows that humans do have the capability to distinguish gait patterns from other patterns. However, these approaches were not automated. Recently, Niyogi and Adelson [3] distinguished different walkers by extracting their spatiotemporal gait patterns obtained from curve-fitting 'snakes'. Cunado *et al.* [4] developed a technique which considers legs as an interlinked penduli and use phase-weighted Fourier magnitude spectra as the extracted feature to recognise different people. Little and Boyd [5] use frequency and phase features from optical flow information to recognise different people by their gait.

IEE Proceedings online no. 19990187
DOI: 10.1049/ip-vis:19990187

Paper first received 17th September 1998 and in revised form 19th January 1999

The authors are with the Image, Speech and Intelligent Systems Research Group, Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK

P.S. Huang will be with the Department of Electrical Engineering, Chung Cheng Institute of Technology, Tashi, Taoyuan, Taiwan 33509, Republic of China from June 1999

Using the human shape and its temporal changes during walking, Murase and Sakai [6] proposed a template matching method which uses the parametric eigenspace representation, as applied in face recognition [7], to recognise different human gait. For recognising people by their gait, this appears to offer greater potential compared with the 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 [8].

Automatic face recognition has been investigated and conducted on various aspects in psychophysics, neurosciences and engineering over the past 20 years. Two survey papers by Samal and Iyengar [9], and Chellapa *et al.* [10] discuss a variety of approaches ranging from Karhunen–Loève expansion [11, 7], feature matching [12, 13], to neural networks [14] in automatic face recognition, using different sources such as video, profile and range imagery. Face image representations based on PCA have been used successfully for various face recognition systems described in [11, 7]. However, PCA based on the global covariance matrix of the full set of image data is not sensitive to the inherent class structure in the data.

To increase the discriminatory power of various facial features, Etemad and Chellappa [8] use linear discriminant analysis, also called canonical analysis which can be used to optimise the class separability of different face classes and improve classification performance. The features are obtained by maximising the between-class variations while minimising the within-class variations. Unfortunately this approach has a high computational cost, and as such was only tested with small images. We call this approach 'canonical space transformation' (CST).

Previously, we proposed a statistical approach [15] which combined eigenspace transformation (EST) with canonical space transformation (CST) based on canonical analysis (CA) for feature extraction of spatial templates to recognise humans by their gait. This statistical approach has been also successfully applied in image retrieval [16] and face recognition [17]. This approach can be used to reduce data dimensionality and to optimise 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 a low-dimensional canonical space. A walking sequence becomes a trajectory in this new space and the distinction between different gaits would appear potentially clearer by dimensionality reduction and class separation.

In this paper we incorporate temporal information from optical-flow changes between two consecutive spatial templates into temporal templates which represent distribution of velocity magnitudes in each pixel and use them for gait recognition. The combined approach [15] of EST and CST is then used for feature extraction. Recognition is achieved in the canonical space by selecting the minimum *accumulated distance* of test template sequences to the centroids of training subjects. Note that the meaning of *temporal templates* used in this paper is different from the same term proposed by Davis and Bobick [18] where they use motion-energy images (MEIs) and motion-history images (MHIs). Unlike our temporal templates computed from optical flow changes between two consecutive images, MEIs and MHIs are actually binary cumulative motion images and temporal-history motion images (at each pixel) through an entire sequence.

Fig. 1 gives an overview of our system showing the steps in training and test procedures.Fig. 1*a* describes the steps in the training procedure and Fig. 1*b* shows how to recognise a gait sequence in the test procedure. The data processing for template projection in Fig. 1*b* is shown in Fig. 2. Details are explained in the following Sections.

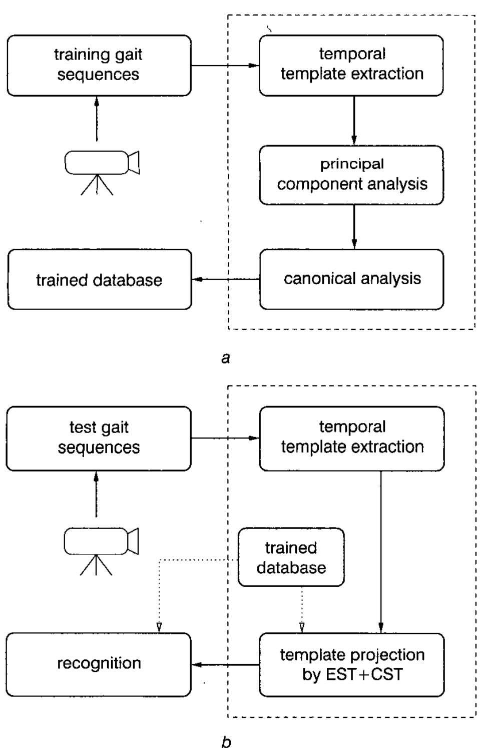

Fig. 1 *Block diagrams of training and test*
a Training diagram
b Test diagram

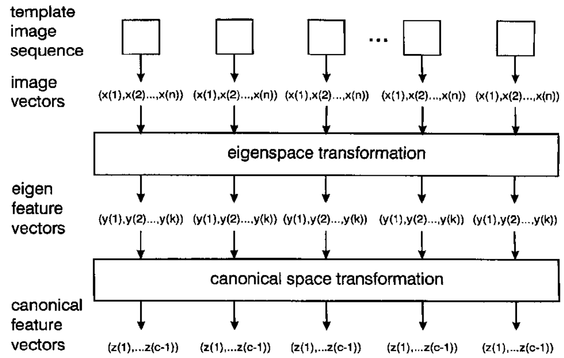

Fig. 2 *Template projection by EST + CST*

2 Preprocessing

We empirically show that recognising humans by gait can be achieved by applying statistical analysis to the temporal patterns of individual subjects. Before training and recognition, each gait sequence is converted into a sequence of temporal templates which are used for recognition. To extract the temporal templates, the centroid and the window containing the silhouette of each template in the original image are used to bound the region for the calculation of optical flow. The extraction processes of spatial and temporal templates are described in the following.

2.1 Extraction of spatial templates

To extract the spatial templates, we choose the preliminary process from Murase's approach [6] in which the silhouette is fitted in a 64×64 image template by normalising its position and size with constant aspect ratio. Naturally, to isolate the human silhouette, we can simply subtract the background from each image. However, simple subtraction can result in noisy regions inside the silhouette area introduced by clothes worn by the subject and by background pixels. To avoid this, a binary image is obtained by region growing [19] followed by subtraction. Fig. 3*a* shows

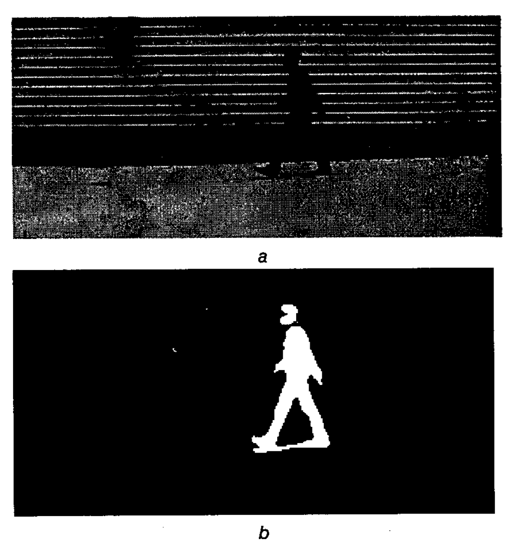

Fig. 3 *Sample images of walking subject*
a Walking subject
b Thresholded version

an image from a gait sequence and Fig. 3*b* is a binary (thresholded) version. The centroid and silhouette window of each template in the original image can be obtained simultaneously from the distribution of segmented pixels and they are used later for the extraction of temporal templates. Sample spatial templates from a gait sequence are illustrated in Fig. 4.

2.2 Extraction of temporal templates

To extract the temporal templates, Little and Boyd's technique [5] which is based on the algorithm of Bulthoff *et al.* [20] is used to generate optical flow fields between two consecutive images. Instead of isolating the moving figure manually, as in [5], we use the information of centroids and silhouette windows from the extraction of spatial templates [15] to extract each temporal template which contains the flow within a moving window. This can reduce the error associated with manual labeling and align the flow fields of the subject with respect to the silhouette centroids.

Unlike other methods, Little and Boyd [5] used dense optical flow fields, generated by minimising the sum of absolute differences between image patches [20]. However, this algorithm is sensitive to any brightness change caused by reflections, shadows. Therefore, the images are first processed by computing the logarithm of brightness, so 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 searches for the displacement of each pixel among a limited set of discrete displacements by minimising the sum of absolute differences between 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 two images are switched. For a correct match, the results will agree. To remove invalid matches, the results at each point in the first image are compared with the result at the corresponding point in the second. The second point should match 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 subpixel displacements but only integer values are used here. In effect, the minimum displacement is 1.0 pixels per frame; points that are assigned nonzero displacements form a set of *moving points*.

Three kinds of temporal templates are generated and they are the u-flow templates which are the horizontal components of flow, the v-flow templates which are the vertical components of flow and $|(u,v)|$-flow templates which are the magnitudes of (u,v). Examples are shown in Fig. 5 where the top row of images are u-flow templates, the middle row are v-flow templates and the bottom row are $|(u,v)|$-flow templates from a gait sequence. For display purposes, stationary pixels are represented by grey-value 0, moving points are shaded. We choose $|(u,v)|$-flow templates which combine two different templates, u-flow and v-flow templates, as features to distinguish different gaits. The comparison of three kinds of temporal features is also shown in the experimental results.

3 Training and projection

Spatial templates are used only to define the bounding regions of temporal templates, not for training. After the extraction of temporal templates, all temporal templates from training sequences are used directly for training by

Fig. 4 *Sample spatial templates*

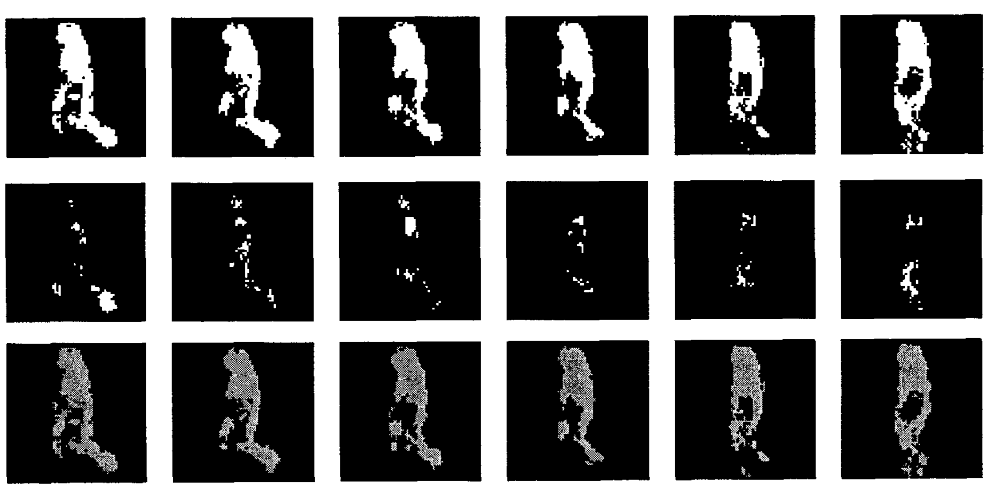

Fig. 5 *Sample*
Top: u-flow templates
Middle: v-flow templates
Bottom: $|(u, v)|$-flow templates

EST and CST. They will be individually projected into the canonical space for further recognition after a training stage. Fig. 2 illustrates the projection steps that generate feature vectors by eigenspace transformation and canonical space transformation for each sequence of template images. Adopted from previous work [15], we use the transformation which combined EST and CST for feature extraction. Template images in the high-dimensional image space are converted to a low-dimensional eigenspace using EST. Obtained vectors thus are further projected to a smaller canonical space using CST. Recognition is accomplished in the canonical space. Patently, the reduced dimensionality results in a concomitant decrease in computation cost.

Given c classes for training and each class represents a temporal template sequence of a single person. $\mathbf{x}_{i,j}$ is the jth template in class i, j is 70 to 80 frames (covering about four walking cycles) in the experiments. Multiple gait sequences of each subject can be added for training without altering the method. N_i is the number of templates in ith class, the total number of training templates is $N_T = N_1 + N_2 + \cdots + N_c$. This training set is represented by

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

where each sample $\mathbf{x}_{i,j}$ is an template image with n pixels ($n = 64 \times 64 = 4096$ in this paper). The mean $\mathbf{m}_\mathbf{x}$ of full image set can be given by

$$\mathbf{m}_\mathbf{x} = \frac{1}{N_T} \sum_{i=1}^{c} \sum_{j=1}^{N_i} \mathbf{x}_{i,j} \tag{2}$$

Here, each pixel value represents the velocity of optical flow for its corresponding position.

3.1 Eigenspace transformation

EST has been used in face recognition [11, 7, 21–23] and to represent [24] and recognise [6] human gait. Basically it is used to reduce the dimensionality of an input space by mapping the data from a correlated high-dimensional space to an uncorrelated low-dimensional space whilst maintaining the minimum mean-square error for the information loss. EST uses the eigenvalues and eigenvectors generated by the data covariance matrix to rotate the original data coordinates along the direction of maximum variance. Let Σ be a $n \times n$ matrix and represented by

$$\Sigma = \frac{1}{N_T} \sum_{i=1}^{c} \sum_{j=1}^{N_i} (\mathbf{x}_{i,j} - \mathbf{m}_\mathbf{x})(\mathbf{x}_{i,j} - \mathbf{m}_\mathbf{x})^T \tag{3}$$

If the rank of matrix Σ is K, the K nonzero eigenvalues of Σ, $\lambda_1, \ldots, \lambda_K$, and associated eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_K$ satisfy the fundamental eigenvalue relationship

$$\lambda_i \mathbf{e}_i = \Sigma \mathbf{e}_i \quad i = 1, \ldots, K \tag{4}$$

However, the computation of eqn. 4 is intractable for normal image size n. Based on *singular value decomposition* theory [25], we can compute another matrix $\tilde{\Sigma}$ instead, that is

$$\tilde{\Sigma} = \frac{1}{N_T} \sum_{i=1}^{c} \sum_{j=1}^{N_i} (\mathbf{x}_{i,j} - \mathbf{m}_\mathbf{x})^T (\mathbf{x}_{i,j} - \mathbf{m}_\mathbf{x}) \tag{5}$$

in which the size of matrix $\tilde{\Sigma}$ is $N_T \times N_T$ and is much smaller than $n \times n$ in practical problems (i.e. $\tilde{\Sigma}$ is a 432×432 matrix as opposed to Σ which is 4096×4096). Suppose that the matrix $\tilde{\Sigma}$ has nonzero eigenvalues $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_K$ and associated eigenvectors $\tilde{e}_1, \ldots, \tilde{e}_K$ which are related to those in Σ by

$$\begin{cases} \lambda_i = \tilde{\lambda} \\ \mathbf{e}_i = \tilde{\lambda}_i^{-1/2} \mathbf{X} \tilde{\mathbf{e}}_i \end{cases} \tag{6}$$

where $i = 1, \ldots, K$ and $\mathbf{X}$ is represented by

$$\mathbf{X} = [\mathbf{x}_{1,1} - \mathbf{m}_\mathbf{x}, \ldots, \mathbf{x}_{1,N_1} - \mathbf{m}_\mathbf{x}, \ldots, \mathbf{x}_{c,N_c} - \mathbf{m}_\mathbf{x}] \tag{7}$$

According to the theory of PCA, each image can be approximated by taking only the $k < K$ largest eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \cdots |\lambda_k|$ 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

$$\mathbf{y}_{i,j} = [\mathbf{e}_1, \ldots, \mathbf{e}_k]^T \mathbf{x}_{i,j} \tag{8}$$

where $i = 1, \ldots, c$ and $j = 1, \ldots, N_c$. The size of $\mathbf{y}_{i,j}$ is 110 in the experiments and the size of $\mathbf{e}_i$ equals the template size (4096).

We call this matrix $[\mathbf{e}_1, \ldots, \mathbf{e}_k]^T$ the *eigenspace transformation matrix.*

3.2 Canonical space transformation (CST)

Based on the theory of canonical analysis [26], CST is presented as follows. Suppose that $\{\Phi_1, \Phi_2, \ldots, \Phi_c\}$ represents the classes of transformed vectors by eigenspace transformation and $\boldsymbol{y}_{i,j}$ is the jth vector in class i. The mean vector of the entire set is given by

$$\mathbf{m}_\mathbf{y} = \frac{1}{N_T} \sum_{i=1}^{c} \sum_{j=1}^{N_i} \mathbf{y}_{i,j} \tag{9}$$

and the mean vector of the ith class is represented by

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{y_{i,j} \in \Phi_i} \mathbf{y}_{i,j} \tag{10}$$

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

$$\mathbf{S_w} = \frac{1}{N_T} \sum_{i=1}^{c} \sum_{y_{i,j} \in \Phi_i} (\mathbf{y}_{i,j} - \mathbf{m}_i)(\mathbf{y}_{i,j} - \mathbf{m}_i)^T$$

$$\mathbf{S_b} = \frac{1}{N_T} \sum_{i=1}^{c} N_i (\mathbf{m}_i - \mathbf{m}_\mathbf{y})(\mathbf{m}_i - \mathbf{m}_\mathbf{y})^T$$

where the sizes of $\mathbf{S_w}$ and $\mathbf{S_b}$ are 110×110 in this paper. To maximise distances between different classes and minimise distances within each class, the objective is to minimise $\mathbf{S_w}$ and maximise $\mathbf{S_b}$ simultaneously. That is to maximise the criterion function known as the *generalised Fisher linear discriminant function* [26] and given by

$$J(\mathbf{W}) = Tr\{(\mathbf{W}\mathbf{S_w}\mathbf{W}^T)^{-1}(\mathbf{W}\mathbf{S_b}\mathbf{W}^T)\} \tag{11}$$

in which matrix $\mathbf{W}$ represents the eigenvectors in each column which maximise $J(\mathbf{W})$. According to [26], the maximisation of eqn. 11 results in the generalised eigenvalue equation

$$\mathbf{S_w}^{-1}\mathbf{S_b}\mathbf{w}_i^* = \lambda_i \mathbf{w}_i^* \tag{12}$$

where λ_i and $\mathbf{w}_i$* are the eigenvalue and eigenvector of $\mathbf{S_w}^{-1}$ $\mathbf{S_b}$. The class separability can be maximised by solving this equation. After eqn. 12 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

$$\mathbf{z}_{i,j} = [\mathbf{v}_1, \ldots, \mathbf{v}_{c-1}]^T \mathbf{y}_{i,j} \tag{13}$$

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 (the size of $\mathbf{z}_{i,j}$ equals five in the paper). We call this orthogonal basis $[\mathbf{v}_1, \ldots, \mathbf{v}_{c-1}]^T$ the 'canonical space transformation matrix'. By merging eqn. 8 and eqn. 13, each image can be projected into one point in the new $(c-1)$-dimensional space by

$$\mathbf{z}_{i,j} = [\mathbf{v}_1, \ldots, \mathbf{v}_{c-1}]^T [\mathbf{e}_1, \ldots, \mathbf{e}_k]^T \mathbf{x}_{i,j} \tag{14}$$

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

$$\mathbf{C}(i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{z}_{i,j} \tag{15}$$

4 Gait recognition

Let a test gait sequence be $\mathbf{g}(t)$, in which $t = 1, \ldots, T$. Before recognition, temporal templates are extracted from this test sequence and projected into a trained canonical space by eqn. 14 yielding one vector sequence after projection,

$$\mathbf{h}(t) = (v_1(t), \ldots, v_{c-1}(t))$$

in which $\mathbf{h}(t)$ represent feature vectors of temporal templates. To recognise a human walking sequence from a trained database in canonical space, the accumulated distance to each centroid is used. This eliminates matching problems caused by velocity changes and phase shifts in spatio-temporal correlation. The accumulated distance between the test sequence $\mathbf{h}(t)$ and c centroids $\mathbf{C}(i)$ in which $i = 1, \ldots, c$ is

$$d_c^2(i) = \sum_{t=1}^{T} \|\mathbf{h}(t) - \mathbf{C}(i)\|^2 \tag{16}$$

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

5 Experimental results

The sample human gait data came from the Visual Computing Group, University of California, San Diego. There are six people and seven sequences of each. One walking sequence is selected from each person as the training sequence and remaining 36 sequences served as test sequences. Each sequence was preprocessed and converted into a temporal template sequence as described in Section 2 before training and projection.

5.1 Eigenspace transformation

For simplicity, we only discuss the results of PCA training for $|(u,v)|$-flow templates. The PCA training of u-flow and v-flow templates are accomplished in the same manner. Fig. 6*a* shows the logarithmic magnitude of each eigenvalue in the eigenspace after six sequences of training $|(u,v)|$-flow templates are trained by PCA. Fig. 6*b* shows their accumulated variances. According to the best result in the experiments, we choose the first 110 eigenvalues which accumulate 90% of the total variance and their corresponding eigenvectors as the eigenspace transformation matrix.

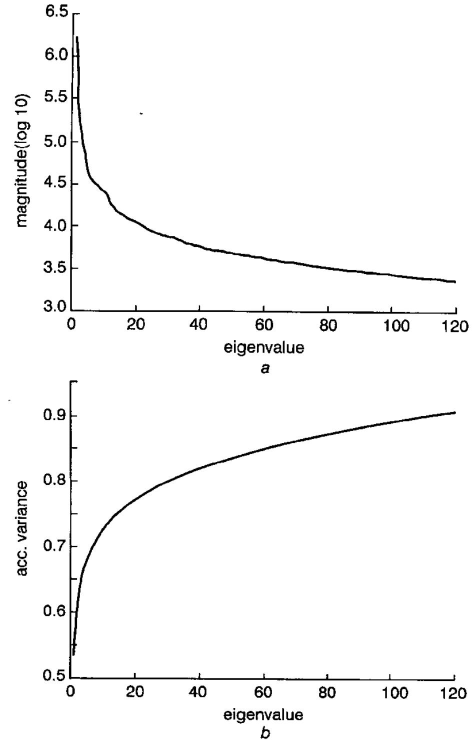

Fig. 6 *Eigenvalues in eigenspace of $|(u,\ v)|$-flow templates*
a Logarithmic magnitudes
b Accumulated variances

Each temporal template is then converted into one point in this eigenspace spanned by the basis of 110 eigenvectors.

Thus, each temporal template can be represented by the linear combination of those 110 principal eigenvectors. Fig. 7 shows the mean image of training templates and first five eigenvectors calculated from the training $|(u,\ v)|$-flow templates which have been normalised for display. Each template is then projected from a matrix to a vector with 110 coefficients which is much smaller than the original (64×64)-dimensional image space. Figs. 8–11 show the distributions of training sequences in different spaces. Fig. 8 shows the distribution of six training sequences in eigenspace of $|(u,\ v)|$-flow templates. It reveals that six training gait sequences appear as six periodic trajectories in the eigenspace.

5.2 Canonical space transformation

After PCA, canonical analysis is further applied to these 110-dimensional vectors converted from temporal templates. Fig. 12 shows the eigenvalues in the canonical space and that there are five nonzero eigenvalues after CA. Thus, the five eigenvectors associated with nonzero eigenvalues are used as the transformation matrix of canonical space and each 110-dimensional vector is projected into a five-dimensional canonical space. Actual recognition is achieved in this space and size of matching is greatly reduced from 64×64 to five for each template pair.

Figs. 9 show the distribution of six training sequences in canonical space of $|(u,v)|$-flow templates. For comparison, the distributions in canonical spaces of u-flow templates and v-flow templates are shown in Figs. 10 and 11. For visualisation, we only show the first three dimensions spanned by the first three principal eigenvectors. As we

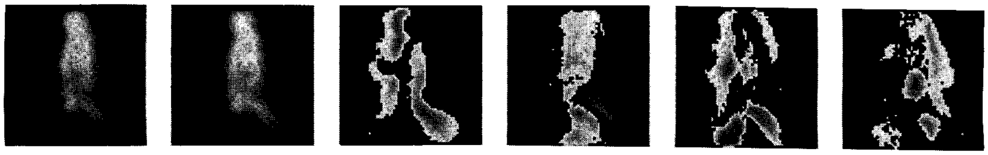

Fig. 7 *Mean vector and five eigenvectors of |(u, v)|-flow templates*
Left to right: mean, e1, e2, e3, e4, e5

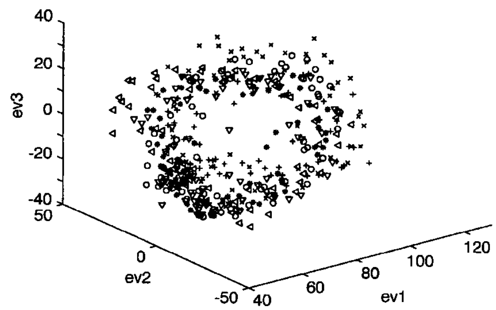

Fig. 8 *Distribution of training sequences in eigenspace of |(u, v)|-flow templates*

○ person 1 △ person 4
* person 2 × person 5
+ person 3 ▽ person 6

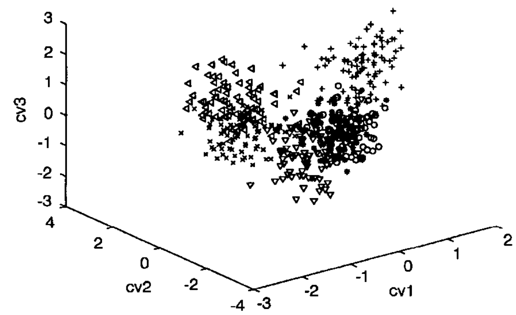

Fig. 11 *Distribution of training sequences in canonical space of v-flow templates*

○ person 1 △ person 4
* person 2 × person 5
+ person 3 ▽ person 6

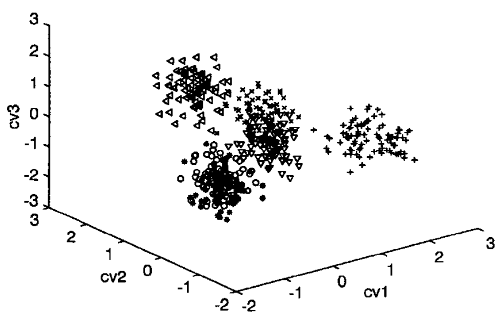

Fig. 9 *Distribution of training sequences in canonical space of |(u, v)|-flow templates*

○ person 1 △ person 4
* person 2 × person 5
+ person 3 ▽ person 6

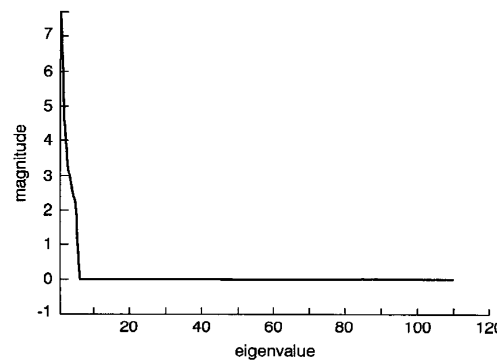

Fig. 12 *Eigenvalues in canonical space of |(u, v)|-flow templates*

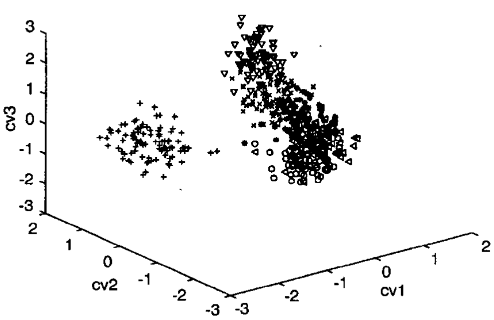

Fig. 10 *Distribution of training sequences in canonical space of u-flow templates*

○ person 1 △ person 4
* person 2 × person 5
+ person 3 ▽ person 6

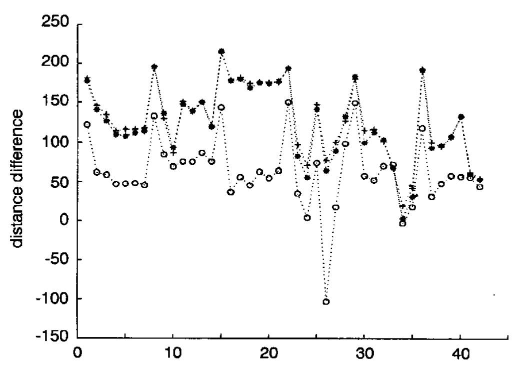

Fig. 13 *Distance measures of 42 sequences in canonical space*

* |(u, v)|-flow templates
+ u-flow templates
○ v-flow templates

can see, the class separability in the canonical space is much better than in the eigenspace.

5.3 Evaluation

In the canonical space, we use eqn. 16 as the distance measure to discriminate between gaits. For comparison, linear rescaling [27] has been applied to each vector to set the average of each data set to zero and to normalise the standard deviation. The distance measures for the recognition using our approach for three kinds of temporal templates is shown in Fig. 13 which shows the differences of accumulated distance between minimum and second minimum matches using three different temporal features. Negative values represent misclassifications. Different peak values accrue from unequal length of gait sequences. Here, the v-flow template has lower distance and hence poorest discriminatory ability. 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 the middle row of Fig. 5. Vertical movements of gait usually have smaller changes than horizontal movements, thus have less discriminatory power in distinguishing different gaits.

To compare with the performance of EST [6] we also calculated the recognition rate in eigenspace of $|(u,v)|$-flow templates. Spatio-temporal distance [6] is used to recognise an input gait sequence in the eigenspace and is given by

$$d_e^2(i) = \min_{\delta} \sum_{t=1}^{T} \|\mathbf{u}(t) - \mathbf{y}_{i,t+\delta}\|^2 \tag{17}$$

in which $\mathbf{u}(t)$ is the projected sequence of $\mathbf{g}(t)$ in the eigenspace and δ represents the phase difference and is an integer. To match a test sequence $\mathbf{u}(t)$ to a training sequence in the eigenspace can be accomplished by choosing i which minimises $d_e^2\ (i)$. The comparison of recognition performance using five different approaches is shown in Table 1.

Clearly, the feature vectors generated by the combination of EST and CST yield the best recognition rates among five approaches. Although promising results have been shown, further evaluation on a larger database is needed.

5.4 Influence of different training samples and eigenvalues

To evaluate the recognition effects in canonical space using different training samples (walking cycles) and eigenvalues, we conducted four tests which used 18, 36, 54 and 72 templates corresponding to one, two, three and four walking cycles from each training sequence for training. Furthermore, In each test, we choose seven different accumulated variances ranging from 65 to 95% achieved by different number of eigenvalues (associated to eigenvectors). The comparison of recognition performance is shown in Table 2 and Fig. 14. The results reveal that the best performance is accomplished by using over 90% accumulated variance of eigenvalues and four walking cycles of training samples from each subject. Furthermore, the recognition performance is improved by increasing the number of training samples and accumulated variances (number of eigenvectors). Thus, results would appear to confirm sensitivity to the number of learned samples and implies that in a more extended analysis, care must be taken to include sufficient samples in the training database.

Table 1: Recognition using different approaches

	Method	Recognition rate, %
(1)	Little & Boyd's	95.2
(2)	EST in $|(u, v)|$-flow templates	92.7
(3)	EST + CST in v-flow templates	95.2
(4)	EST + CST in u-flow templates	100
(5)	EST + CST in $|(u, v)|$-flow templates	100

Table 2: Recognition rates using different training samples and eigenvalues

Training samples	Accumulated variance of eigenvalues, % 65	70	75	80	85	90	95
1 cycle	21.4	45.2	69.0	83.3	85.7	83.3	85.7
2 cycles	19.0	78.6	90.5	95.2	95.2	92.9	95.2
3 cycles	14.3	81.0	90.5	95.2	97.6	97.6	97.6
4 cycles	69.0	90.5	95.2	97.6	97.6	100	100

6 Discussions and conclusions

Although gait patterns may be different depending on age, sex, clothing, etc., we only consider the gait as a discriminatory feature between different subjects rather than considering any effect introduced by the subject. Such factors will doubtless be of concern in later work, but our concern here is more the basic nature of gait, as described by standard measures. As such, we have used the $|(u,v)|$-flow templates as features for gait recognition. Although the u-flow templates achieves largely similar (or even slightly better) performance than the $|(u,v)|$-flow templates in the experiments, $|(u,v)|$-flow templates (which include the magnitude of u-flow and v-flow templates) have potentially more information than u-flow templates. Naturally, the flow technique is sensitive to changes in frame rate and in speed of the subject. However, this is not a issue here where the sequences are of exactly the same type, but for other data the technique would need to be modified via reformulation to compensate for these factors.

EST greatly reduces the dimensionality of each template by projecting each template from a highly correlated high-dimensional space to an uncorrelated low-dimensional

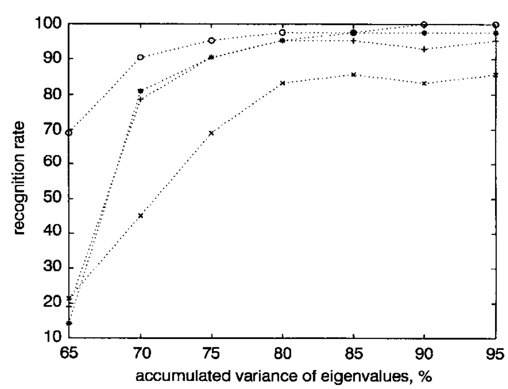

Fig. 14 *Recognition rates using different training samples and eigenvalues*

○ four walking cycles
* three walking cycles
+ two walking cycles
× one walking cycle

space. Based on the global scatter matrix of the training set, EST is not sensitive to class structure in the data. Thus, EST is only used to reduce the template dimension in the first stage. CST improves the class separability by maximising the between-class variations while minimising the within-class variations. However, the singularity problem appears when CST is applied directly to templates and the template size is greater than the number of training samples. Therefore CST is adopted to process the projected vectors in the eigenspace in the second stage.

Our previous approach which combined EST with CST for feature extraction has been extended with new motion data. EST and CST can be used to reduce data dimensionality and to optimise the class separability of different classes simultaneously, greatly improving the performance of eigenspace approach. Not using the feature of spatial templates developed in previous work, we adopt the feature of temporal templates. This new feature incorporates temporal information into each template. The performance comparison for the eigenspace approach, Little and Boyd's approach and our method in gait recognition shows that the combination of EST and CST is better than the other two approaches. By these results we have revealed the recognition scheme of human gait by statistical analysis; recognising humans by gait can be achieved by analysing the temporal patterns of individual subjects.

We are working on extending the test environment. In addition to testing on a larger database, the extended features by the combination of different features will be also investigated. For extended feature vectors, it has been also suggested in [13] that orthogonal feature sets should be chosen to reduce the variance of a final match measure, which might prove to be of benefit. Future work will also concentrate on looking for more precise and robust features, while aiming to develop the technique still further.

7 Acknowledgments

Dr. Jeffrey Boyd at the Visual Computing Laboratory, University of California, San Diego, USA provided gait data and gave invaluable advice.

8 References

1 CUTTING, J.E., and KOZLOWSKI, L.T.: 'Recognizing friends by their walk: gait perception without familiarity cues', *Bull. Psychonom. Soc.*, 1977, **9**, (5), pp. 353–356
2 JOHANSSON, G.: 'Visual perception of biological motion and a model for its analysis', *Percept. Psychophys.*, 1973, **14**, (2), pp. 201–211
3 NIYOGI, S.A., and ADELSON, E.H.: 'Analysis and recognizing walking figures in XYT'. proceedings of IEEE conference on *Computer vision and pattern recognition*, June 1994, Seattle, WA, USA, pp. 469–474
4 CUNADO, D., NIXON, M.S., and CARTER, J.N.: 'Using gait as a biometric, via phase-weighted magnitude spectra' Proceedings of First international conference on *Audio- and video-based biometric person authentication*, AVBPA '97, March 1997, Crans-Montana, Switzerland, pp. 95–102
5 LITTLE, J., and BOYD, J.: 'Recognizing people by their gait: the shape of motion', *MIT Press J. - Videre*, 1998, **1**, (2), pp. 1–32
6 MURASE, H., and SAKAI, R.: 'Moving object recognition in eigenspace representation: gait analysis and lip reading', *Pattern Recognit. Lett.*, 1996, **Vol. 17**, pp. 155–162
7 TURK, M., and PENTLAND, A.: 'Eigenfaces for recognition', *J. Cogn. Neurosci.*, 1991, **3**, pp. 71–86
8 ETEMAD, K., and CHELLAPPA, R.: 'Discriminant analysis for recognition of human face images', *J. Opt. Soc. Am. A, Opt. Image Sci. Vis.*, 1997, **14**, (8), pp. 1724–1733
9 SAMAL, A., and IYENGAR, P.: 'Automatic recognition and analysis of human faces and facial expressions: A survey', *Pattern Recognit.*, 1992, **25**, pp. 65–77
10 CHELLAPPA, R., WILSON, C.L., and SIROHEY, S.: 'Human and machine recognition of faces: a survey', *Proc. IEEE*, 1995, **83**, (5), pp. 705–740
11 KIRBY, M., and SIROVICH, L.: 'Application of the Karhunen – Loève procedure for the characterization of human faces', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1990, **12**, (1), pp. 103–108
12 BRUNELLI, R., POGGIO, T.: 'Face recognition: Features versus templates', *IEEE Trans. Pattern. Anal. Mach. Intell.*, 1993, **15**, (10), pp. 1042–1052
13 JIA, X., and NIXON, M.S.: 'Extending the feature vector for automatic face recognition', *IEEE Trans. Pattern. Anal. Mach. Intell.*, 1995, **17**, (12), pp. 1167–1176
14 VALENTIN, D., ABDI, H., O'TOOLE, A.J., and COTTRELL, G.W.: 'Connectionist models of face processing: a survey', *Pattern Recognit.*, 1994, **27**, pp. 1209–1230
15 HUANG, P.S., HARRIS, C.J., and NIXON, M.S.: 'Canonical space representation for recognizing humans by gait and face'. Proceedings of IEEE southwest symposium on *Image analysis and interpretation*, April 1998, Tucson, AZ, USA, pp. 180–185
16 SWETS, D.L., and WENG, J.J.: 'Using discriminant eigenfeatures for image retrieval', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **18**, (8), pp. 831–836
17 BELHUMEUR, P.N., HESPANHA, J.P., and KRIEGMAN, D.J.: 'Eigenfaces vs. fisherfaces: recognition using class specific linear projection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 711–720
18 DAVIS, J.W., and BOBICK, A.F.: 'The representation and recognition of human movement using temporal templates'. Proceedings of IEEE conference on *Computer vision and pattern recognition*, June 1997, San Juan, Puerto Rico, pp. 928–934
19 DUBUISSON, M.-P., and JAIN, A.K.: 'Contour extraction of moving objects in complex outdoor scenes', *Int. J. Comput. Vision*, 1995, **14**, (6), pp. 83–105
20 BULTHOFF, H., LITTLE, J., and POGGIO, T.: 'A parallel algorithm for real-time computation of optical flow', *Nature*, February 1989, **337**, pp. 549–553
21 MOGHADDAM, B., and PENTLAND, A.: 'Probabilistic visual learning for object representation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 696–710
22 PENG, H., and ZHANG, D.: 'Dual eigenspace method for human face recognition', *Electron. Lett.*, 1997, **33**, (4), pp. 283–284
23 YUEN, P.C., DAI, D.Q., and FENG, G.C.: 'Wavelet-based PCA for human face recognition'. Proceedings of IEEE southwest symposium on *Image analysis and interpretation*, April 1998, Tucson, AZ, USA, pp. 223–228
24 BLACK, M.J., YACOOB, Y., JEPSON, A.D., and FLEET, D.J.: 'Learning parameterized models of image motion'. Proceedings of IEEE conference on *Computer vision and pattern recognition*, June 1997, San Juan, Puerto Rico, pp. 561–567
25 MURAKAMI, H., and KUMAR, V.: 'Efficient calculation of primary images from a set of images', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1982, **PAMI-4**, (5), pp. 511–515
26 FUKUNAGA, K.: 'Introduction to statistical pattern recognition', (Academic Press, 1990, 2nd edn.)
27 BISHOP, C.M.: 'Neural networks for pattern recognition', (Oxford University Press, New York, 1996)