Human Carrying Status in Visual Surveillance

Dacheng Tao[1], Xuelong Li[1], Xindong Wu[2], and Stephen J. Maybank[1]

1. School of Computer Science and Information Systems, Birkbeck, University of London.
2. Department of Computer Science, University of Vermont.

{dacheng, xuelong, sjmaybank}@dcs.bbk.ac.uk; xwu@cs.uvm.edu

Abstract

A person's gait changes when he or she is carrying an object such as a bag, suitcase or rucksack. As a result, human identification and tracking are made more difficult because the averaged gait image is too simple to represent the carrying status. Therefore, in this paper we first introduce a set of Gabor based human gait appearance models, because Gabor functions are similar to the receptive field profiles in the mammalian cortical simple cells. The very high dimensionality of the feature space makes training difficult. In order to solve this problem we propose a general tensor discriminant analysis (GTDA), which seamlessly incorporates the object (Gabor based human gait appearance model) structure information as a natural constraint. GTDA differs from the previous tensor based discriminant analysis methods in that the training converges. Existing methods fail to converge in the training stage. This makes them unsuitable for practical tasks.

Experiments are carried out on the USF baseline data set to recognize a human's ID from the gait silhouette. The proposed Gabor gait incorporated with GTDA is demonstrated to significantly outperform the existing appearance-based methods.

1. Introduction

Gait recognition [13][32][29] is important in visual surveillance because it is possible to identify a person at a distance and without the person being aware that he or she is under observation. Among factors affecting human gait, the *carrying* status [12], i.e. whether or not the person is carrying an object (such as a bag, suitcase or rucksack) plays a key role. E.g. in the recent London incidents (Fig. 1), cameras recorded people's gait when they were *carrying* rucksacks and these sequences were important evidence for agencies fighting against international terrorism.

Figure 1: (http://www.smh.com.au) Human gait when *carrying* rucksacks - captured by passive visual surveillance in connection with the recent incidents in London (Luton).

Human brains are constantly analysing motion information from external world. The human gait not only produces a distinctive moving silhouette but also reflects the walker's physical situation and psychological state. The original research on gait was entirely from the medical perspective, e.g., Murray [23] used gait to classify pathologically abnormal patients into several groups for suitable treatments.

In computer vision research, human gait recognition is promising because of its good potential for human identification in visual surveillance. It is often difficult to deduce a person's age or gender from a static picture taken at a distance, but a dynamic image sequence may reveal this information [3]. Appearance-based models have been employed in gait research [24]. These utilize either the entire silhouette [21][5][18][11][12] or the most discriminant parts [9][17][31], such as the torso.

For a recognition problem, features and classification stand as two convex mirrors each reflecting and amplifying the other. Herein, we aim to solve two issues, i.e., representation and classifier design, for gait recognition under carrying status.

Features for object representation: we develop a set of *Gabor based gait* features to represent the gait silhouette sequences; and

Classification: we develop a *general tensor discriminant analysis* (GTDA), motivated by the successes of tensor methods in signal processing [16], computer vision [30][26], and pattern recognition

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

[33][27]. Unlike some previous efforts on tensor based discriminant analysis, our algorithm *converges* during the training procedure with a differential scatter discriminant criterion (DSDC). The proposed methods can also be utilized in many other fields, such as face recognition [4][15], texture classification, and image retrieval. The contributions of the paper are:

1. Introduce the Gabor functions to represent the gait image and develop three new methods to utilize the Gabor functions for gait recognition;
2. Deduce DSDC and generalize this criterion for tensor based discriminant analysis to reduce the *ill-posed* problem in linear discriminant analysis (LDA) and achieve convergence in the training phase.

The rest of the paper is organized as follows. Section 2 develops three methods to utilize Gabor functions for gait representation. In Section 3, LDA is briefly reviewed, and GTDA is proposed with DSDC. Section 4 shows that our proposed method significantly outperforms the existing algorithms for gait recognition and conclusions are drawn in Section 5.

2. Gabor Gait Representation

Marcelja [22] and Daugman [8] modeled the visual cortex by Gabor functions, because they are similar to the receptive field profiles in the mammalian cortical simple cells. Daugman [8] developed the 2D Gabor functions, a series of local spatial bandpass filters, which achieve the resolution limit specified by the Heisenberg uncertainty principle and have the characteristics of spatial locality, orientation selectivity, and frequency selectivity. A Gabor (wavelet, kernel, or filter) function is the product of an elliptical Gaussian envelope and a complex plane wave, defined as:

$$\psi_{s,d}(x,y) = \psi_{\bar{k}}(\bar{x}) = \psi_{s,d}(\bar{x}) = \frac{\|\bar{k}\|}{\delta^2} \cdot e^{-\frac{\|\bar{k}\|^2 \cdot \|\bar{x}\|^2}{2\delta^2}} \cdot \left[e^{i\bar{k}\cdot\bar{x}} - e^{-\delta^2/2} \right] \quad (1)$$

where $\bar{x} = (x, y)$ is the variable in a spatial domain and $\bar{k}$ is the frequency vector, which determines the scale and the orientation of Gabor functions, $\bar{k} = k_s e^{i\phi_d}$, where $\phi_d = \pi d/8$, with $0 \le d \le 7$ and $k_s = k_{\max}/f^s$, $k_{\max} = \pi/2$, $f = 2$, $0 \le s \le 4$. The parameters used here follow [20]. The number of oscillations under the Gaussian envelope function is determined by $\delta = 2\pi$. The term $\exp(-\sigma^2/2)$ is subtracted in order to make the kernel DC–free, thus making the filter values insensitive to changes in illumination.

Each particular Gabor function can be used as a mother function to generate a whole family of Gabor functions by scaling and rotating the wave vector. Examples of the Gabor functions (the real part) used in our experiments are presented in Fig 2. We use Gabor functions with five different scales and eight different orientations, making a total of forty Gabor functions.

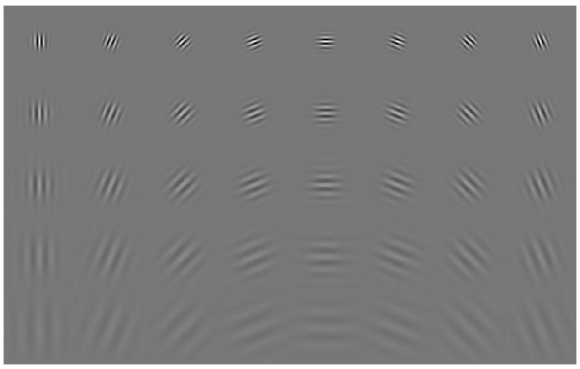

Figure 2: The real part Gabor functions with five different scales (vertical axis) and eight different orientations (horizontal axis).

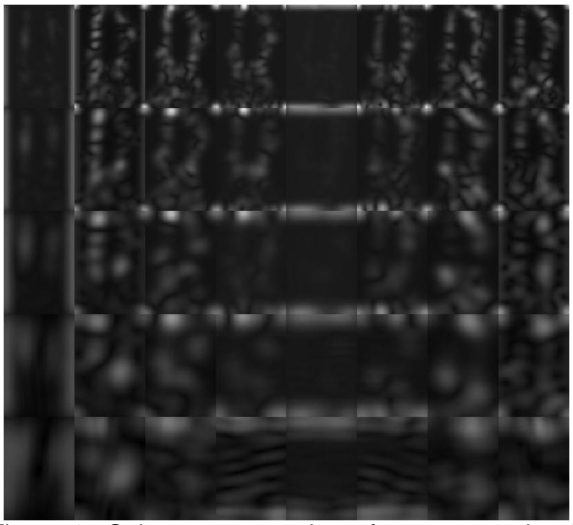

Figure 3: Gabor representation of an averaged gait image.

The Gabor function representation of a gait image can be obtained by convolving the Gabor functions with the gait image. This gives forty filtered images, which represent the original gait image. Examples of the filtered images are given in Fig 3. The method is identical to the face representation [20] using Gabor functions. Although this method is powerful in face recognition, it is less effective for gait recognition, as shown by the experiments in Section 4.2.

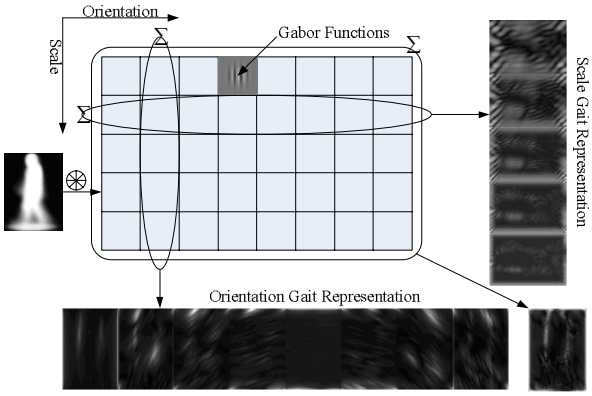

Figure 4: Three new methods for averaged gait image representation using Gabor functions: the orientation gait representation (OGR), the scale gait representation (SGR), and the total Gabor gait representation (TGR).

Due to the poor performance of the existing Gabor representation of gait images, we introduce three new representations. These are the orientation gait representation (OGR), the scale gait representation (SGR), and the total Gabor gait representation (TGR). Fig. 4 shows the procedure for OGR, SGR, and TGR generation.

OGR is given by summing the images obtained at different filter scales, when the orientation of the filter is fixed:

$$\begin{aligned} \mathrm{OGR}(x,y) &= \sum_s I(x,y) * \psi_{s,d}(x,y) \\ &= I(x,y) * \sum_s \psi_{s,d}(x,y) \end{aligned} \tag{2}$$

where $I(x,y)$ is the averaged gait image, $\psi_{s,d}(x,y)$ is the Gabor function defined in (1), and $\mathrm{OGR}(x,y)$ is the output of the OGR method for representation. Therefore, we have eight different outputs to represent the original gait image.

SGR is given by summing the images obtained at different filter orientations, when the scale is fixed,

$$\begin{aligned} \mathrm{SGR}(x,y) &= \sum_d I(x,y) * \psi_{s,d}(x,y) \\ &= I(x,y) * \sum_d \psi_{s,d}(x,y). \end{aligned} \tag{3}$$

Therefore, we have five different outputs to represent the original gait image.

TGR is given by summing the images obtained at all orientations and scales,

$$\begin{aligned} \mathrm{TGR}(x,y) &= \sum_s \sum_d I(x,y) * \psi_{s,d}(x,y) \\ &= I(x,y) * \sum_s \sum_d \psi_{s,d}(x,y). \end{aligned} \tag{4}$$

Therefore, we have one output to represent the original gait image.

3. General Tensor Discriminant Analysis

Human gait images are usually represented by second–order tensors. The Gabor representation contributes an additional dimension giving a third–order tensor representation of gait. In this paper, we do not regard the scale parameter and the orientation parameter as two modes but one. Traditionally, such high order tensors must be scanned into long vectors to comply with the input requirements of conventional discriminant analysis [10]. However, during vectorization, the intrinsic structure information [1] is lost. The number of components in the vector is much larger than the size of the training (gallery) set. The dimensionality of the feature space is much larger than the number of training samples. This often leads to an ill-posed classification problem.

We argue that the current poor recognition rate in gait recognition is due to the ill-posed nature of the classification. Focusing on this problem, our method, named general tensor discriminant analysis (GTDA), is a representation and classification scheme for high–order tensor data, such as images and video. GTDA represents these high–order tensor data in their original format without destroying their intrinsic structures. In this section, we first review LDA and then introduce the modified discriminant criterion.

We use the following notation. Bold uppercase symbols such as **X**, **Y**, **Z** represent tensor objects; normal uppercase symbols such as X, Y, Z represent matrices; italic lowercase symbols such as x, y, z represent vectors; normal lowercase symbols such as a, b, c represent scale numbers; and i, j, k, l represent the indices in a vector, a matrix, or a tensor.

3.1. Linear Discriminant Analysis (LDA)

LDA finds a projection of the $x_{i;j}$, which is optimal for separating the different classes of the $x_{i;j} \in R^N$, where i is the class number, $1 \le i \le c$, and j is the sample ID in the i^{th} class with $1 \le j \le n_i$. The projection U is chosen to maximize the ratio between the trace of the between-class scatter matrix S_b and the trace of the within-class scatter matrix S_w:

$$U = \arg\max_U \frac{\mathrm{Tr}\left(U^T S_b U\right)}{\mathrm{Tr}\left(U^T S_w U\right)}. \tag{5}$$

[1] In this paper we use the term "structure information" of an image to mean information about the relative positions of pixels or regions in the image.

The scatter matrices S_b and S_w are defined by:

$$\begin{cases} S_b = \frac{1}{n}\sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T \\ S_w = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i} (x_{i;j} - m_i)(x_{i;j} - m_i)^T \end{cases} \quad (6)$$

where $n = \sum_{i=1}^{c} n_i$ is the size of the training set, n_i is the number of training samples in the i^{th} class, $m = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i} x_{i;j}$ is the mean vector of the total training set, $m_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{i;j}$ is the mean vector for the individual class C_i, and $x_{i;j}$ is the j^{th} sample in the i^{th} individual class C_i.

The optimal projection matrix U can be computed from the leading eigenvectors of $S_w^{-1}S_b$. In many computer vision applications, the dimensionality N of the feature space is much larger than the size of the training set (i.e., $N \gg n$). Since the rank of S_w is at most $n-c$, S_w is singular if N is large, and it is more difficult to construct U. This is known as the Small Sample Size (SSS) problem. In this paper, we use the PCA plus LDA method [11] to avoid the degeneration of S_w. The dimensionality of the LDA space is fixed as $c-1$ for gait recognition. More information can be found from [11] about the method for gait recognition.

3.2. Differential Scatter Discriminant Criterion

The Differential Scatter Discriminant Criterion (DSDC) [10] is defined by:

$$U = \arg\max_U \left(\mathrm{Tr}\left(U^T S_b U\right) - \zeta \,\mathrm{Tr}\left(U^T S_w U\right)\right), \quad (7)$$

where ζ is a tuning parameter; $U \in R^{N\times N^*}$ ($N^* \ll N$), constrained by $U^T U = I$, is the optimal classification projection matrix; and S_b, S_w are defined in (6).

According to [10] (pp.446-447) and [27], the optimal solution in (7) is equal to the optimal solution in (5), if we set ζ as the Lagrange multiplier. In [27], $\zeta = \lambda_{\max}\left(S_w^{-1}S_b\right)$, which is the maximum eigenvalue of $S_w^{-1}S_b$, to achieve the optimal solution, if we extract only one feature (U which is a vector). When we want to extract t features simultaneously, we can approximately estimate ζ as $\sum_{i=1}^{t}\lambda_i\left(S_w^{-1}S_b\right)$, where $\lambda_i|_{i=1}^{t}$ are the first t maximum eigenvalues of $S_w^{-1}S_b$. From [10] (pp. 446-447), it is not difficult to obtain that the optimal ζ in (7) is $\mathrm{Tr}\left(U_{opt}^T S_b U_{opt}\right)/\mathrm{Tr}\left(U_{opt}^T S_w U_{opt}\right)$[2]. An accurate solution of (7) can be obtained by the alternating projection method.

In real-world applications, because the distribution of the testing set diverges from the distribution of the training set, we manually set ζ to achieve better prediction results.

3.3. General Tensor Discriminant Analysis

Because $\mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$, based on (7), we have:

$$U = \arg\max_U \left(\mathrm{Tr}\left(U^T (S_b - \zeta S_w) U\right)\right). \quad (8)$$

Because $\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA)$, based on (8), we have:

$$U = \arg\max_U \left(\mathrm{Tr}\left((S_b - \zeta S_w) UU^T\right)\right). \quad (9)$$

By replacing S_b and S_w according to (6), based on (9) we have:

$$\begin{aligned} U &= \arg\max_U \left(\mathrm{Tr}\left((S_b - \zeta S_w) UU^T\right)\right) \\ &= \arg\max_U \mathrm{Tr}\left(\left[\begin{array}{l} \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T \\ -\zeta \sum_{i=1}^{c}\sum_{j=1}^{n_i} (x_{i;j} - m_i)(x_{i;j} - m_i)^T \end{array}\right] UU^T\right) \\ &= \arg\max_U \left\| \left[\begin{array}{l} \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T \\ -\zeta \sum_{i=1}^{c}\sum_{j=1}^{n_i} (x_{i;j} - m_i)(x_{i;j} - m_i)^T \end{array}\right] UU^T \right\|_{Fro} \end{aligned} \quad (10)$$

where $\|\cdot\|_{Fro}$ is the Frobenius norm and the projection matrix $U \in R^{N\times N^*}$ ($N^* < N$) constrained by $U^T U = I$.

Based on (10), we can analogously obtain the general tensor discriminant criterion naturally as:

---

[2] The derivative of $\mathrm{Tr}\left(U^T S_b U\right) - \zeta\,\mathrm{Tr}\left(U^T S_w U\right)$ with U is given by $S_b U - \zeta S_w U$. To obtain the optimal solution of (7), we need to set $S_b U - \zeta S_w U$ as 0 (as we have a strict condition here, i.e., $(S_b - \zeta S_w)u_k = 0$, $\forall u_k \in U$, u_k is a column vector in U). Consequently, we have $\mathrm{Tr}\left(U_{opt}^T S_b U_{opt}\right) = \zeta\,\mathrm{Tr}\left(U_{opt}^T S_w U_{opt}\right)$, where U_{opt} is the optimal solution of (7).

$$U_l\,|_{l=1}^{M} = \arg\max_{U_l\,|_{l=1}^{M}} \left\| \left(\begin{array}{l} \sum_{i=1}^{c} n_i (\mathbf{M}_i - \mathbf{M})(\mathbf{M}_i - \mathbf{M})^{T_l} \\ -\zeta_l \sum_{i=1}^{c} \sum_{j=1}^{n_i} (\mathbf{X}_{i;j} - \mathbf{M}_i)(\mathbf{X}_{i;j} - \mathbf{M}_i)^{T_l} \end{array} \right) \prod_{l=1}^{M} \otimes U_l U_l^T \right\|_{Fro}, \quad (11)$$

where $\mathbf{X}_{i;j}$ denotes the j^{th} training sample (tensor) in the i^{th} individual class C_i; $\mathbf{M}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{X}_{i;j}$ is the i^{th} class mean tensor; $\mathbf{M} = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i} \mathbf{X}_{i;j} = \frac{1}{n}\sum_{i=1}^{c} n_i \mathbf{M}_i$ is the total mean tensor of all training tensors; U_l denotes the l^{th} mode projection matrix for decomposition in the training procedure; and $\otimes$ denotes the tensor product. Moreover, $\mathbf{X}_{i;j}\,|_{1\le i\le c}^{1\le j\le n_i}$, $\mathbf{M}_i\,|_{i=1}^{c}$, and $\mathbf{M}$ are all M^{th}–order tensors that lie in the $R^{N_1\times N_2\times \ldots N_M}$ space; and the notation $\cdot^{T_l}$ means the l^{th} transpose/permutation of the tensor by rearranging the l^{th} mode to the 1^{st} mode. As pointed in Section 3.2., ζ_l is the l^{th} Lagrange/tuning parameter.

The problem defined in (11) is the so-called *General Tensor Descriminant Analysis* (GTDA). GTDA has no closed form solution and we choose the alternating projection method to obtain the optimal projection matrices $U_l\,|_{l=1}^{M}$, i.e., we can optimize U_d in GTDA (11) with the given $U_l\,|_{1\le l\le M}^{l\ne d}$ iteratively until convergence. In the t^{th} iteration, with the given $U_{l;(t-1)}\,|_{1\le l\le M}^{l\ne d}$ in the $(t\text{-}1)^{\text{th}}$ iteration and the matrix unfolding[3] operation defined in [16], we can obtain the optimal d^{th} mode projection matrix $U_{d;(t)}$ by maximizing:

$$\left\| \left(\begin{array}{l} \sum_{i=1}^{c} n_i \left(\bar{M}_{i;(d)} - \bar{M}_{(d)}\right)\left(\bar{M}_{i;(d)} - \bar{M}_{(d)}\right)^T \\ -\zeta_l \sum_{i=1}^{c}\sum_{j=1}^{n_i} \left(\bar{X}_{i;j;(d)} - \bar{M}_{i;(d)}\right)\left(\bar{X}_{i;j;(d)} - \bar{M}_{i;(d)}\right)^T \end{array} \right) U_{d;(t)} U_{d;(t)}^T \right\|_{Fro} \quad (12)$$

where $\bar{M}_{i;(d)}$ means the d^{th} mode matrix unfolding on $\mathbf{M}_i \prod_{l=1;l\ne d}^{M} \otimes U_{l;(t-1)}$; $\bar{X}_{i;j;(d)}$ is the d^{th} mode matrix unfolding on $\mathbf{X}_{i;j} \prod_{l=1;l\ne d}^{M} \otimes U_{l;(t-1)}$; and $\bar{M}_{(d)}$ is the d^{th} mode matrix unfolding on $\mathbf{M} \prod_{l=1;l\ne d}^{M} \otimes U_{l;(t-1)}$. To reduce the complexity of the training procedure, in this paper we set ζ_l as the maximum eigenvalue of

$$\left[\frac{\sum_{i=1}^{c} n_i \left(\bar{M}_{i;(l)} - \bar{M}_{(l)}\right)\left(\bar{M}_{i;(l)} - \bar{M}_{(l)}\right)^T}{\sum_{i=1}^{c}\sum_{j=1}^{n_i} \left(\bar{X}_{i;j;(l)} - \bar{M}_{i;(l)}\right)\left(\bar{X}_{i;j;(l)} - \bar{M}_{i;(l)}\right)^T} \right] \quad (13)$$

where $\left[\frac{A}{B}\right]$ means $B^{-1}A$.

With this setting, we can achieve a better gait recognition rate.

Furthermore, for the first iteration in optimization, we can choose the $U_l\,|_{1\le l\le M}^{l\ne d}$ to be random matrices or $\mathbf{1}_{N_{l^*}\times N_l}$. The algorithm to obtain the optimal solution of GTDA is given by Table 1.

According to the algorithm described in Table 1, we obtain $U_l\,|_{l=1}^{M} \in R^{N_{l^*}\times N_l}$ $N_{l^*} < N_l$ iteratively. For GTDA, we use the projected tensor $\mathbf{Y} = \mathbf{X}\prod_{l=1}^{M}\times_l U_l$ to represent the original general tensor $\mathbf{X}$. For recognition, the prototype $\mathbf{X}^p$ for each individual class in the database and the test tensor $\mathbf{X}^t$ to be classified are projected by projection matrices $U_l\,|_{l=1}^{M}$ onto $\mathbf{Y}^p$ and $\mathbf{Y}^t$, respectively. The test tensor class is found by minimizing the distance $\varepsilon = \left\|\mathbf{Y}^p - \mathbf{Y}^t\right\|$ over p.

Unlike the existing tensor extension of discriminant analysis [32], the training stage of GTDA converges. The following method can check the convergence of GTDA. In step 5 of Table 1, we check the convergence through $\left\|\left|U_{l(t)}U_{l(t-1)}^T\right| - I\right\| \le \varepsilon$ for all $1 \le l \le M$, where ε is small. If $\left|U_{l(t)}U_{l(t-1)}^T\right| = I$, the calculated projection direction in the t^{th} or the current iteration is equivalent to the $(\text{t}-1)^{\text{th}}$ or the last iteration.

3.4. Why Use a Tensor Representation?

To vectorize a tensor into a vector makes it harder to keep track of the information in spatial constraints. For example, two 4-neighbor connected pixels in an image may be separated hugely from each other after a vectorization.

To better characterize or classify natural data, a scheme should preserve as many as possible of the original constraints. When the training samples are limited, these constraints help to give reasonable solutions to classification problems. Take the strategy

[3] The d^{th} mode matricizing or matrix unfolding of an M^{th} order tensor $\mathbf{X}$ are tensors in R^{N_d} obtained by keeping index d fixed and varying the other indices. Therefore, the d^{th} mode matrix unfolding $X_{(d)}$ is in $R^{N_d \times \prod_{1\le l\le M}^{l\ne d} N_d}$.

in the Gaussian distribution estimation as an example[4]: when the training samples embedded in a high dimensional space are limited, we always add some constraints to the covariance matrix, such as requiring the covariance matrix to be a diagonal matrix.

Table 1. An iterative algorithm to solve GTDA.

Input: Training tensors $\mathbf{X}_i \,|_{i=1}^{n} \in R^{N_1 \times N_2 \times \ldots N_M}$ and the dimensionality of the output tensors $\mathbf{Y}_i \in R^{N_{1^*} \times N_{2^*} \times \ldots N_{M^*}}$.

Output: The projection matrix $U_l \,|_{l=1}^{M} \in R^{N_{l^*} \times N_l}$ constrained by $U_l^T U_l = I$ and the output tensors $\mathbf{Y}_i \,|_{i=1}^{n} \in R^{N_{1^*} \times N_{2^*} \times \ldots N_{M^*}}$.

Initialization: $U_l = \mathbf{1}_{N_{l^*} \times N_l}$.

Step 1. For $t = 1$ to T {

Step 2. For $d = 1$ to M {

Step 3. Calculate ζ_l according to (13);
Optimize $U_{d;(t)}$ with the given $U_{l;(t-1)} \,|_{1 \le l \le M}^{l \ne d}$ by maximizing (12) by conducting the singular value decomposition on:

Step 4.
$$\begin{pmatrix} \sum_{i=1}^{c} n_i \left(\bar{M}_{i;(d)} - \bar{M}_{(d)} \right) \left(\bar{M}_{i;(d)} - \bar{M}_{(d)} \right)^T \\ -\zeta_l \sum_{i=1}^{c} \sum_{j=1}^{n_i} \left(\bar{X}_{i;j;(d)} - \bar{M}_{i;(d)} \right) \left(\bar{X}_{i;j;(d)} - \bar{M}_{i;(d)} \right)^T \end{pmatrix}.$$

}//For loop in Step 2.

Step 5. Convergence checking: the training stage of GTDA converges, if $\left\| U_{l(t)} U_{l(t-1)}^T \right| - I \|_{Fro} \le \varepsilon$ for all $1 \le l \le M$.

}//For loop in Step 1.

Step 6. Calculate output tensors $\mathbf{Y}_i = \mathbf{X}_i \prod_{l=1}^{M} \times_l U_l$.

The tensor representation helps to reduce the number of parameters needed to model the data. For example, when a tensor X has the size $n_1 \times n_2$, we need to estimate the projection matrix U with the size $(n_1 \times n_2) \times n$ for traditional LDA (n is the number of selected features), but we only need to estimate the projection matrices U_1 with the size $n_1 \times n$ and U_2 with the size $n_2 \times n$ in GTDA. Furthermore, the estimation procedures for the 1st mode projection matrix U_1 and the 2nd mode projection matrix U_2 are independent. Consequently, the number of the parameters in GTDA is much less than that of LDA.

[4] Constraints in TR1DA are justified by the form of the data. However, constraints in the example are ad hoc.

Therefore, the under sample problem can be reduced through GTDA because GTDA could be deemed as a constrained LDA according to:

$$\mathrm{vec}(X \otimes U_1 \otimes U_2) = (U_2 \otimes U_1)\,\mathrm{vec}(X). \quad (14)$$

3.5. Convergence Issue for GTDA

To describe the convergence of GTDA, we need to define the following functions based on (12) and (13), respectively:

$$f_U(d,t) = f_U\left(U_{d;(t)}; \zeta_{l;(t)}, U_{l;(t)} \,|_{l=1}^{d-1}, U_{l;(t-1)} \,|_{l=d+1}^{M}\right) = \left\| \begin{pmatrix} \sum_{i=1}^{c} n_i \left(\bar{M}_{i;(d)} - \bar{M}_{(d)} \right) \left(\bar{M}_{i;(d)} - \bar{M}_{(d)} \right)^T \\ -\zeta_{l;(t)} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \left(\bar{X}_{i;j;(d)} - \bar{M}_{i;(d)} \right) \left(\bar{X}_{i;j;(d)} - \bar{M}_{i;(d)} \right)^T \end{pmatrix} U_{d;(t)} U_{d;(t)}^T \right\|_{Fro} \quad (15)$$

$$f_\zeta(d,t) = f_\zeta\left(U_{l;(t)} \,|_{l=1}^{d-1}, U_{l;(t-1)} \,|_{l=d+1}^{M}\right) = \zeta_{d;(t)} = \lambda\left(\left[\frac{\sum_{i=1}^{c} n_i \left(\bar{M}_{i;(d)} - \bar{M}_{(d)} \right) \left(\bar{M}_{i;(d)} - \bar{M}_{(d)} \right)^T}{\sum_{i=1}^{c} \sum_{j=1}^{n_i} \left(\bar{X}_{i;j;(d)} - \bar{M}_{i;(d)} \right) \left(\bar{X}_{i;j;(d)} - \bar{M}_{i;(d)} \right)^T} \right] \right). \quad (16)$$

where $\lambda(X)$ means the maximum eigenvalue of X.

With (15) and (16), we can define:

$$g_U(d,t) = \max_{U_{l;(t)} \in R^{N_d^* \times N_d}} f_U(d,t), \quad (17)$$

$$g_\zeta(d,t) = \max_{\zeta_d} f_\zeta(d,t). \quad (18)$$

Formally, the alternating projection method never decreases $F_U(d,t)$ and $F_\zeta(d,t)$ with the increasing t and d, because $g_U(d,t)$ and $g_\zeta(d,t)$ are convex.

4. Experimental Results

This section first briefly describes the USF HumanID gait database [24] (gallery and probe data sets), and the performance of the baseline system [24] is reported. We then compare the performances of our schemes with [11] and [24].

4.1. HumanID Gait Database

We carried out all of our experiments upon the USF HumanID outdoor gait database, which has been built and widely utilized for vision–based gait recognition. It consists of 1,870 sequences from 122 subjects. For each of the subjects, there are the following covariates: change in viewpoints (**L**eft or **R**ight), change in shoe type (**A** or **B**), change in walking surface (**G**rass or **C**oncrete), change in carrying condition (carrying a

Brie**F**case or **N**o **B**riefcase), and change in elapsed time (**M**ay or **N**ovember) between sequences being compared. All these covariates are very important for different aspects/applications; among them, the carrying status is particularly important. In the USF gait gallery [24] there are three pre-designed experiments for algorithm comparisons when the carrying status is positive. For algorithm training, the database provides a gallery with the following covariates: grass, shoe type A, right camera, and no briefcase.

4.2. Comparison Experiments

Table 2. Rank 1/5 experimental results for human gait recognition. The left group with columns H, I, and J is the rank 1 performance and the right group with columns H, I, and J is the rank 5 performance.

Rank 1/5	**H**	**I**	**J**	**H**	**I**	**J**
Baseline	61	57	36	85	78	62
LDA[1]	63	59	54	90	81	79
LDA[2]	62	60	57	89	86	77
GTDA[1]	65	59	29	92	79	59
GTDA[2]	90	**88**	58	**97**	95	79
GTDA[3]	**92**	84	**66**	**97**	**98**	**85**
GTDA[4]	**92**	78	59	95	95	81

We compare the performances of our schemes with [11] and [24]. The dissimilarity measure for the gait recognition is the same as with [24],

$$\begin{aligned}&\mathrm{Dist}\left(AS_P^{Method}, AS_G^{Method}\right)\\&=\mathrm{Median}_{i=1}^{N_P}\left(\min_{j=1}^{N_G}\left\|AS_P^{Method}(i)-AS_G^{Method}(j)\right\|\right)\end{aligned} \tag{19}$$

where $AS_P^{Method}(i)|_{i=1}^{N_P}$ is the i^{th} projected averaged silhouette or AS in the probe data and $AS_G^{Method}(j)|_{j=1}^{N_G}$ is the j^{th} projected AS in the gallery. The equation measures the dissimilarity, which is the median value of the N_p outputs from the minimum Euclidean distance between the averaged silhouettes from the i^{th} AS of the probe and all AS of the gallery.

Focusing on appearance-based models in human gait recognition, we implemented all proposed methods together with some important conventional methods, such as LDA[1] (LDA with real gait) [11], LDA[2] (LDA with gait feature fusion) [11], and the baseline algorithm [24]. To give a compact report about our comparison experiments, we build a set of abbreviations to denote our different schemes, namely, GTDA[1] (GTDA using the conventional Gabor method), GTDA[2] (GTDA using OGR), GTDA[3] (GTDA using SGR), and GTDA[4] (GTDA using TGR). As introduced in Section 1, we mainly focus on the carrying status issue, and it corresponds to tasks **H**, **I** and **J** [24] in these tables.

Table 2 reports all comparing experiments in clear manners. The bold numbers show "which scheme gives the best results for this problem". In the table, the first three rows give out the performances of the baseline algorithms, LDA[1] and LDA[2], respectively; while the performances of the new schemes upon the same gallery set and probe set are fully reported on all the comparison experiments. As shown in Table 2, the recognition rate is significantly improved in terms of both rank 1 and rank 5 evaluations by using our GTDA[3].

4.3. Convergence Examination

In this part, we mainly study the convergence of the proposed GTDA. From Figure 5, it can be inferred that only 3 to 5 iterations are usually required to achieve GTDA convergence. In contrast, the traditional 2DLDA cannot converge during the training procedure, which can be seen from the first figure in [32].

5. Conclusions

This paper proposes effective methods to solve the *carrying status* problem in visual surveillance. We first introduced a multi-resolution representation of human gait through three different Gabor filter based multi-resolution representations for the averaged silhouette sequences. To reduce the *ill-posed* problem, we also developed the *general tensor discriminant analysis* (GTDA), which converges well during the training stage. From *un-convergence* to *convergence*, an huge improvement was achieved in comparison with the previous tensor-based learning algorithms. By these means, upon the USF baseline platform, we reported *state-of-the-art* results on the *carrying status* issue of gait recognition.

6. References

[1] C. Ben – Abdelkader, R. Cutler, and L. Davis, "Person Identification using Automatic Height and Stride Estimation," ICPR, vol. 4, pp. 377–380, Quebec, 2002.

[2] A. Bobick and A. Johnson, "Gait Recognition using Static Activity–specific Parameters," CVPR, vol. 1, pp. 423–430, Kauai, HI, 2001.

[3] J. E. Boyd, "Synchronization of Oscillations for Machine Perception of Gaits," CVIU, vol. 96, no. 1, pp. 35–59, 2004.

[4] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a Similarity Metric Discriminatively, with Application to Face Verification," CVPR, vol. 1, pp. 539-546, 2005.

[5] R. T. Collins, R. Bross, and J. Shi, "Silhouette – based Human Identification from Body Shape and Gait," FG, pp. 351–356, Washington DC, 2002.
[6] D. Cunado, M. Nixon, and J. Carter, "Automatic Extraction and Description of Human Gait Models for Recognition Purposes," CVIU, vol. 90, no. 1, pp. 1–41, 2003.
[7] R. Cutler and L. Davis, "Robust Periodic Motion and Motion Symmetry Detection," CVPR, pp. 615–622, 2000.
[8] J. G. Daugman, "Two–Dimensional Spectral Analysis of Cortical Receptive Field Profile," *Vision Research*, vol. 20, pp. 847–856, 1980.
[9] J. W. Davis and A. F. Bobick, "The Representation and Recognition of Human Movement using Temporal Templates," CVPR, pp. 928–934, 1997.
[10] K. Fukunaga. Introduction to Statistical Pattern Recognition (2nd ed.). Academic Press, Boston 1990.
[11] J. Han and B. Bhanu, "Statistical Feature Fusion for Gait–Based Human Recognition," CVPR, vol. 2, pp. 842–847, Washington, DC, 2004.
[12] I. Haritaoglu, R. Cutler, D. Harwood, and L. Davis, "Backpack: Detection of People Carrying Objects using Silhouettes," CVIU, vol. 6, no. 3, pp. 2001.
[13] G. Johansson, "Visual Perception of Biological Motion and a Model for its Analysis," *Perception and Psychophysics*, vol. 14, no. 2, pp. 201-211, 1973.
[14] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy–Chowdhury, V. Kruger, and R. Chellappa, "Identification of Humans using Gait," *IEEE Trans. Image Processing*, vol. 13, no. 9, pp. 1163–1173, 2004.
[15] J. Kittler, R. Ghaderi, T. Wideatt, and J. Matas, "Face Verification via Error Correcting Output Codes," IVC, vol. 21(13-14), pp. 1163-1169, 2003.
[16] L. D. Lathauwer, Signal Processing Based on Multilinear Algebra, Ph.D. Thesis, Katholike Universiteit Leuven, 1997.
[17] L. Lee, G. Dalley, and K. Tieu, "Learning Pedestrian Models for Silhouette Refinement," ICCV, vol. 1, pp. 663–670, Nice, France, 2003.
[18] L. Lee and W. E. L. Grimson, "Gait Analysis for Recognition and Classification," FG, pp. 155–162, 2002.
[19] J. J. Little and J. E. Boyd, "Recognizing People by Their Gait: the Shape of Motion," *Videre*, vol. 1, no. 2, pp. 1–32, 1998.
[20] C. Liu, "Gabor–based Kernel PCA with Fractional Power Polynomial Models for Face Recognition," TPAMI, vol. 26, no. 5, pp. 572–581, 2004.
[21] F. Liu and R. W. Picard, "Finding Periodicity in Space and Time," ICCV, pp. 376–383, Bombay, India, 1998.
[22] S. Marcelja, "Mathematical Description of the Responses of Simple Cortical Cells," *Journal of the Optical Society of America*, vol. 70, no. 11, pp. 1297–1300, 1980.
[23] M. Murray, A. Drought, and R. Kory, "Walking Pattern of Normal Men," *Journal of Bone and Joint Surgery*, vol. 46–A, no. 2, pp. 335–360, 1964.
[24] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer, "The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis," TPAMI, vol. 27, no. 2, pp. 162–177, 2005.
[25] R. Tanawongsuwan and A. Bobick, "Gait Recognition from Time–normalized Joint–angle Trajectories in the Walking Plane," CVPR, vol. 2, pp. 726–731, 2001.
[26] D. Tao, X. Li, W. Hu, S. J. Maybank, and X. Wu, "Supervised Tensor Learning," *Proc. IEEE Int'l Conf. on Data Mining*, pp. 450-457, November 2005.
[27] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Tensor Rank One Discriminant Analysis," Submitted, 2005.
[28] J. Tenenbaum and W. Freeman, "Separating Style and Content," NIPS, vol. 10, pp. 662–668, 1997.
[29] N. Troje, "Decomposing Biological Motion: A Framework for Analysis and Synthesis of Human Gait Patterns," *Journal of Vision*, vol. 2, no. 5, pp. 371–387, 2002.
[30] M. A. O. Vasilescu, D. Terzopoulos, "Multilinear Subspace Analysis for Image Ensembles," CVPR, vol.2, pp. 93–99, Madison, WI, 2003.
[31] G. Veres, L. Gordon, J. Carter, and M. Nixon, "What Image Information is Important in Silhouette–based Gait Recognition?," CVPR, vol. 2, pp. 776–782, 2004.
[32] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification," TPAMI, pp. 1505-1518, December 2003.
[33] J. Ye, R. Janardan, and Q. Li, "Two–Dimensional Linear Discriminant Analysis," NIPS, pp. 1,569–1,576, 2005.

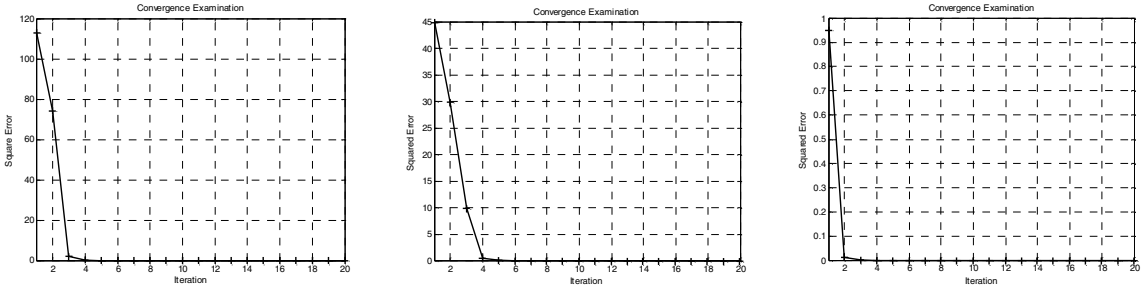

Figure 5: Three orientation convergence examination curves in the GTDA4 (GTDA using SGR) training procedure