

BOOSTING LDA WITH REGULARIZATION ON MPCA FEATURES FOR GAIT RECOGNITION

Haiping Lu, K.N. Plataniotis and A.N. Venetsanopoulos

The Edward S. Rogers Sr. Department of Electrical and Computer Engineering
University of Toronto, M5S 3G4, Canada
{haiping, kostas, anv}@comm.toronto.edu

ABSTRACT

In this paper, we present a boosted linear discriminant analysis (LDA) solution with regularization on features extracted by the multilinear principal component analysis (MPCA) for the gait recognition problem. This work is an extension of a recent LDA-based boosting approach and the MPCA is employed to project tensorial gait samples on a number of discriminative EigenTensorGaits (ETGs) to produce gait feature vectors for the base learners in boosting. This new scheme offers one more way to control the learner weakness while being very computationally efficient. Furthermore, the LDA learners are modified through regularization for protection against overfitting on the gallery set. Promising experimental results obtained on the Gait Challenge data sets indicate that the proposed algorithm is an efficient and effective solution consistently enhancing the gait recognition results on the seven probe sets by MPCA+LDA.

1. INTRODUCTION

The demand for automated person identification system and visual surveillance at a distance is growing considerably in security-sensitive environments such as banks, parking lots, airports and large civic structures, where the verification and identification of subjects is essential. However, in this setting of human recognition at a distance, many conventional biometrics such as fingerprint, face or iris information are not available at high enough resolution for recognition and gait recognition, the identification of individuals through the way they walk, has emerged as a promising solution with the advantages of unobtrusiveness, hard-to-hide, and recognition at a distance [1–3].

Model-based approaches [4, 5] and appearance-based approaches [1, 2] for gait recognition have been studied in the literature, with the latter being more successful. Appearance-based approaches take binary gait silhouette sequences extracted from the raw gait sequences [6] as the input. Gait silhouette sequences are naturally third-order tensors and the three dimensions are the spatial row, column and the temporal modes. To deal directly with these tensor objects in

a very high-dimensional tensor space, traditional linear feature extraction algorithms based on vector input such as the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) first reshape (vectorize) the input tensor objects into vectors in a very high-dimensional space, resulting in high computation and memory demand. Furthermore, the structure and correlation in the original data are broken by this input reshaping and the redundancy and structure in the original data is not fully utilized. This motivated the recent development of multilinear subspace algorithms operating directly on the gait sequences in their tensor representation, rather than their vectorized versions. The multilinear PCA (MPCA) algorithm proposed in [7] aims to determine a multilinear projection that projects the original tensor objects into a lower-dimensional tensor subspace while preserving the variation in the original data. A number of discriminative features in the projected tensor space have been used for gait recognition and good results have been achieved.

However, many factors, such as viewing angles, walking surfaces and shoes, may affect a person's gait and as in the case of face patterns, the gait patterns are expected to be nonlinear and complex. Moreover, the gait data available for training and testing may be captured under different conditions and good generalization is very difficult, as studied in the Gait Challenge problem [2]. There are many methods proposed in the literature to handle complex and nonlinear patterns and the emerging ensemble-based machine learning method called boosting is a very promising one offering good generalization capability through combining a set of (weak) base learners repeatedly trained on weighted training samples [8, 9]. Boosting requires an appropriate weak learner to work, which has restricted its applicability [9, 10] and a recent work in [11] has broken this limitation by proposing a boosting algorithm that works with LDA-style learners, applied on the problem of face recognition. A cross-validation mechanism is employed to weaken the LDA learner and the pairwise class discriminant distribution is introduced for interaction between the booster and the learner.

This paper extends the boosting work in [11] for gait recognition by making use of the recent development of MPCA in

[7]. It should be noted that, to the best of the authors' knowledge, this is the first work combining ensemble-based learning with tensorial subspace solutions. In this novel processing scheme, the MPCA of [7] is first used to generate EigenTensorGaits (ETGs) in a lower-dimensional tensor space and then only a number of discriminative ETGs are selected as the input to the LDA-based booster. This ETG number provides us another way (besides the cross-validation mechanism in [11]) to control the weakness of the LDA learner, and the MPCA feature extractor before the booster greatly reduces the processing cost (in both training and testing) so that the very-high dimensional tensorial gait data can be handled efficiently. In addition, a novel regularization control mechanism is added to improve the generalization as the within-class scatter of gait patterns under various capturing conditions is expected to be larger than that of gait patterns captured under a single condition. In the following section, the proposed algorithm will be discussed in detail. Experimental results and conclusions are presented in Sec. 3 and Sec. 4, respectively.

2. BOOSTING LDA WITH REGULARIZATION ON MPCA FEATURES FOR GAIT RECOGNITION

The block diagram of the proposed method for gait recognition is shown in Figure 1. Input tensorial gait samples are projected on a number of discriminative EigenTensorGaits (ETGs) to obtain gait feature vectors¹ and these vectors are fed into the LDA-based booster for learning and classification. In this section, after the introduction of the notations, the MPCA-based feature extraction from tensorial gait data is briefly reviewed and then the adopted LDA-based boosting solution is presented.

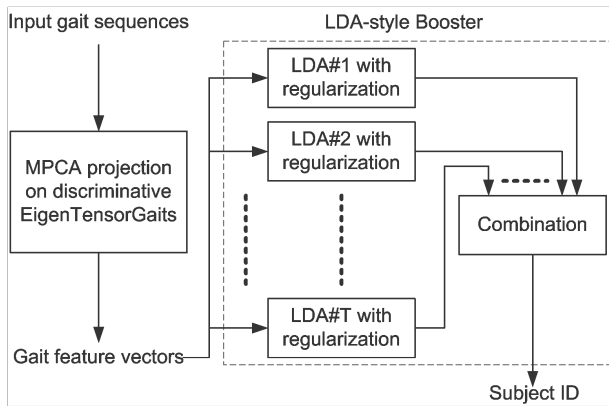


Fig. 1. Illustration of gait recognition through boosting LDA with regularization on MPCA features.

¹The generation of ETGs are not included in the figure and it is briefly described in Sec. 2.2.

2.1. Notations

In this paper, vectors are denoted by lowercase boldface letters, e.g., \mathbf{x} ; matrices by uppercase boldface, e.g., \mathbf{U} ; and tensors by calligraphic letters, e.g., \mathcal{A} . Their elements are denoted with indices in brackets. Indices are denoted by lowercase letters and span the range from 1 to the uppercase letter of the index, e.g., $n = 1, 2, \dots, N$. An N^{th} -order tensor is denoted as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$. It is addressed by N indices i_n , $n = 1, \dots, N$, and each i_n addresses the n -mode of \mathcal{A} . The n -mode product of a tensor \mathcal{A} by a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{U}$, is a tensor with entries: $(\mathcal{A} \times_n \mathbf{U})(i_1, \dots, i_{n-1}, j_n, i_{n+1}, \dots, i_N) = \sum_{i_n} \mathcal{A}(i_1, \dots, i_N) \cdot \mathbf{U}(j_n, i_n)$. A rank-1 tensor \mathcal{A} equals to the outer product of N vectors: $\mathcal{A} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \dots \circ \mathbf{u}^{(N)}$, which means that $\mathcal{A}(i_1, i_2, \dots, i_N) = \mathbf{u}^{(1)}(i_1) \cdot \mathbf{u}^{(2)}(i_2) \cdot \dots \cdot \mathbf{u}^{(N)}(i_N)$ for all values of indices.

2.2. Review of MPCA-based gait feature extraction

In MPCA-based gait feature extraction, each half gait cycle is treated as a gait sample (3rd-order tensor), obtained as in [7]. In a typical gait recognition problem, there are two types of gait data sets: the gallery and the probe [2]. The gallery set contains the set of gait samples with known identities and it is used as the training set. The probe set is the testing set with gait samples of unknown identities to be identified through matching against the gallery set.

The M gallery gait samples $\{\mathcal{X}_1, \dots, \mathcal{X}_M \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}\}$ are used as the input to MPCA. The objective of MPCA is to find a multilinear transformation $\{\tilde{\mathbf{U}}^{(n)} \in \mathbb{R}^{I_n \times P_n}, n = 1, \dots, N\}$ mapping from the original gait tensor space $\mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \dots \otimes \mathbb{R}^{I_N}$ into a tensor subspace $\mathbb{R}^{P_1} \otimes \mathbb{R}^{P_2} \dots \otimes \mathbb{R}^{P_N}$ (with $P_n < I_n$, for $n = 1, \dots, N$): $\mathcal{Y}_m = \mathcal{X}_m \times_1 \tilde{\mathbf{U}}^{(1)T} \times_2 \tilde{\mathbf{U}}^{(2)T} \dots \times_N \tilde{\mathbf{U}}^{(N)T}$, $m = 1, \dots, M$, such that $\Psi_{\mathcal{Y}} = \sum_{m=1}^M \|\mathcal{Y}_m - \bar{\mathcal{Y}}\|_F^2$, the total tensor scatter, is maximized, where $\bar{\mathcal{Y}} = \frac{1}{M} \sum_{m=1}^M \mathcal{Y}_m$ is the mean sample. This problem is solved in an iterative alternating projection manner.

The obtained projection matrices $\{\tilde{\mathbf{U}}^{(n)}, n = 1, \dots, N\}$ can be viewed as $\prod_{n=1}^N P_n$ EigenTensorGaits (ETGs): $\tilde{\mathcal{U}}_{p_1 p_2 \dots p_N} = \tilde{\mathbf{u}}_{p_1}^{(1)} \circ \tilde{\mathbf{u}}_{p_2}^{(2)} \circ \dots \circ \tilde{\mathbf{u}}_{p_N}^{(N)}$, where $\tilde{\mathbf{u}}_{p_n}^{(n)}$ is the p_n^{th} column of $\tilde{\mathbf{U}}^{(n)}$. However, not all of them are useful for recognition and they are selected according to their class discriminability $\Gamma_{p_1 p_2 \dots p_N}$, where $\Gamma_{p_1 p_2 \dots p_N}$ for the eigentensor $\tilde{\mathcal{U}}_{p_1 p_2 \dots p_N}$ is defined as

$$\Gamma_{p_1 p_2 \dots p_N} = \frac{\sum_{c=1}^C N_c \cdot [\bar{\mathcal{Y}}_c(p_1, p_2, \dots, p_N) - \bar{\mathcal{Y}}(p_1, p_2, \dots, p_N)]^2}{\sum_{m=1}^M [\mathcal{Y}_m(p_1, p_2, \dots, p_N) - \bar{\mathcal{Y}}_{c_m}(p_1, p_2, \dots, p_N)]^2}, \quad (1)$$

where C is the number of classes (subjects), M is the total number of gait samples in the gallery set, N_c is the number of gait samples for class (subject) c and c_m is the class label for the m^{th} gallery gait sample \mathcal{X}_m . \mathcal{Y}_m is the feature tensor of \mathcal{X}_m in the projected tensor subspace, the mean fea-

ture tensor $\bar{\mathcal{Y}} = \frac{1}{M} \sum_m \mathcal{Y}_m$ and the class mean feature tensor $\bar{\mathcal{Y}}_c = \frac{1}{N_c} \sum_{m, c_m=c} \mathcal{Y}_m$. For the selection, the entries in \mathcal{Y}_m are arranged into a feature vector \mathbf{y}_m ordered according to $\Gamma_{p_1 p_2 \dots p_N}$ in descending order and only the first H_y entries of \mathbf{y}_m are kept.

2.3. Boosting regularized LDA on MPCA features

The boosting scheme in the proposed approach is summarized in Fig. 2, with details below. It should be noted that the booster proposed here has an important difference with that in [11]. The LDA-style base learners in our booster take $\{\mathbf{y}_m \in \mathbb{R}^{H_y}, m = 1, \dots, M\}$, the gait feature vectors extracted by MPCA, rather than the original data $\{\mathbf{x}_m \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}, m = 1, \dots, M\}$ as in [11]². There are two benefits from the proposed scheme:

1. The gait feature vector dimension H_y , which is the number of discriminative ETGs selected, gives us one more degree (besides R , the number of samples per class used for LDA learners) to control the weakness of the LDA learners. Similar to the case of PCA+LDA, where the performance is often affected by the number of principal components selected for LDA, H_y affects the performance of LDA on the MPCA features as well, as observed in [7]. Therefore, by choosing a H_y that is not optimal for a single LDA learner, the obtained LDA learner is weakened. Of course, the LDA learner cannot be made “too weak” either. Otherwise, the boosting scheme will not work.
2. Using feature vectors of dimension H_y instead of the original data as the booster input is computationally advantageous. Since boosting is an iterative algorithm, the computational cost is about T times of that of a single learner with the same input, both in training and testing. By making the booster to work on lower dimensional features extracted by MPCA, the booster becomes very efficient since it only needs to deal with low-dimensional vectors. For instance, the dimension of the input vectors to the booster is 180 in this paper, which is much smaller than the dimension 17154 for face data in [11] and the original gait data dimension 225280, therefore, the computational cost is reduced significantly.

As in [11], the boosting algorithm in Fig. 2 follows the so-called AdaBoost.M2 algorithm [8]. AdaBoost.M2 is a multi-class extension of the original AdaBoost algorithm (for binary class) [9] and it aims to extend communication between the boosting algorithm and the weak learner by allowing the weak learner to generate more expressive hypotheses (a set of “plausible” labels rather than a single label) indicating a

²Direct application of [11] on the gait recognition problem requires the vectorization of the tensorial input $\{\mathcal{X}_m\}$ to $\{\mathbf{x}_m\}$.

“degree of plausibility”, i.e., a hypothesis h takes a sample \mathbf{y} and a class label c as the inputs and produces a “plausibility” score $h(\mathbf{y}, c) \in [0, 1]$ as the output (rather than a binary output). To achieve this objective, a more sophisticated error measure pseudo-loss ϵ_t is introduced with respect to the so-called mislabel distribution $\mathbf{D}_t(m, c)$ in [8], where t is the boosting step index. A mislabel is a pair (m, c) where m is the index of a training sample and c is an incorrect label associated with the sample m . Let B be the set of all mislabels:

$$B = \{(m, c) : m = 1, \dots, M, c \neq c_m\}. \quad (2)$$

The pseudo-loss ϵ_t of the t^{th} hypothesis h_t with respect to $\mathbf{D}_t(m, c)$ is given by [8]:

$$\epsilon_t = \frac{1}{2} \sum_{(m,c) \in B} \mathbf{D}_t(m, c) (1 - h_t(\mathbf{y}_m, c_m) + h_t(\mathbf{y}_m, c)). \quad (3)$$

Another distribution $\mathbf{d}_t(m)$, named as pseudo sample distribution in [11], is derived from $\mathbf{D}_t(m, c)$ as

$$\mathbf{d}_t(m) = \sum_{c \neq c_m} \mathbf{D}_t(m, c). \quad (4)$$

For the communication between the booster and the learner, the modified “pairwise class discriminant distribution” (PCDD) $\hat{\mathbf{A}}_t \in \mathbb{R}^{C \times C}$ introduced in [11] is employed as

$$\hat{\mathbf{A}}_t(c_a, c_b) = \frac{1}{2} \left(\sum_{c_m=c_a, c_{m_t}=c_b} \mathbf{d}_t(m) + \sum_{c_m=c_b, c_{m_t}=c_a} \mathbf{d}_t(m) \right), \quad (5)$$

where $c_{m_t} = \arg \max_c h_t(\mathbf{y}_m, c)$ and the diagonal of $\hat{\mathbf{A}}_t$ is set to zeros. This version of PCDD results in more independence and diversity between learners, which tends to achieve a low generalization error.

In building the LDA learner, the approach in [11] is adopted with several modifications. Firstly, only R samples per class are used as the input to the LDA learner in order to get weaker but more diverse LDA learners. The first R samples are taken for the first boosting step and the hardest R (with the largest $\mathbf{d}(m)$) samples are selected for subsequent steps. Let $\{\mathbf{y}_s, s = 1, \dots, S\}$ denote the selected samples, where $S = R \times C$. Next, for the between-class scatter matrix \mathbf{S}_B , the pairwise between-class scatter in [12] is used instead of that used in [11] in this paper for its simplicity and easy computation:

$$\mathbf{S}_B = \sum_{c_a=1}^{C-1} \sum_{c_b=c_a+1}^C \hat{\mathbf{A}}_t(c_a, c_b) (\bar{\mathbf{y}}_{c_a} - \bar{\mathbf{y}}_{c_b})(\bar{\mathbf{y}}_{c_a} - \bar{\mathbf{y}}_{c_b})^T, \quad (6)$$

where $\bar{\mathbf{y}}_c = \frac{1}{R} \sum_s^{c_s=c} \mathbf{y}_s$. There is no need to include the priors in \mathbf{S}_B since the number of training samples in each class is the same. Finally, for the within-class scatter matrix, a regularized version of that in [11] is used:

$$\mathbf{S}_W = \sum_s \mathbf{d}(s) (\mathbf{y}_s - \bar{\mathbf{y}}_{c_s})(\mathbf{y}_s - \bar{\mathbf{y}}_{c_s})^T + \eta \cdot \mathbf{I}_{H_y}, \quad (7)$$

Input: The gallery gait feature vectors $\{\mathbf{y}_m, m = 1, \dots, M\}$ with class labels $\mathbf{c} \in \mathbb{R}^M$, the LDA learner described in Sec. 2.3, the number of samples per class for LDA training R , the maximum number of iterations T .

Algorithm:

Initialize $\mathbf{D}_1(m, c) = \frac{1}{M(C-1)}$, $\hat{\mathbf{A}}_1(c_a, c_b) = \frac{1}{C^2}$, $\mathbf{D}_1(m, c_m) = 0$, $\hat{\mathbf{A}}_1(c_a, c_a) = 0$, and the first R samples from each class is selected to form the initial training set $\{\mathbf{y}_s, s = 1, \dots, S\}_1$.

Do for $t=1:T$:

1. Get \mathbf{V}_t from \mathbf{S}_{B_t} and \mathbf{S}_{W_t} constructed from $\{\mathbf{y}_s, s = 1, \dots, S\}_t$ and project $\{\mathbf{y}_m\}$ to $\{\mathbf{z}_m\}$.
2. Get hypothesis $\{h_t(\mathbf{y}_m, c) \in [0, 1]\}$ by applying the nearest mean classifier on $\{\mathbf{z}_m\}$.
3. Calculate ϵ_t , the pseudo-loss of h_t , from (3).
4. Set $\beta_t = \epsilon_t / (1 - \epsilon_t)$.
5. Update \mathbf{D}_t :

$$\mathbf{D}_{t+1}(m, c) = \mathbf{D}_t(m, c) \beta_t^{\frac{1}{2}(1+h_t(\mathbf{y}_m, c_m)-h_t(\mathbf{y}_m, c))},$$

and normalize it:

$$\mathbf{D}_{t+1}(m, c) = \frac{\mathbf{D}_{t+1}(m, c)}{\sum_m \sum_c \mathbf{D}_{t+1}(m, c)}.$$

6. Update $\mathbf{d}_{t+1}(m)$, $\hat{\mathbf{A}}_{t+1}$ and $\{\mathbf{y}_s\}_{t+1}$ accordingly.

Output: The final hypothesis:

$$h_{fin}(\mathbf{y}) = \arg \max_c \sum_{t=1}^T \left(\log \frac{1}{\beta_t} \right) h_t(\mathbf{y}, c)$$

Fig. 2. The pseudo-code implementation of the LDA-based booster.

where η is a regularization parameter to increase the estimated within-class scatter and \mathbf{I}_{H_y} is an identity matrix of size $H_y \times H_y$. The regularization term is added because in the gait recognition problem, the actual within-class scatter of gait sequences captured under various conditions is expected to be greater than the within-class scatter that can be estimated from the gallery set, which is captured under a single condition. With these definitions, the projection \mathbf{V} is then to be solved to maximize the ratio of the between-class scatter to the within-class scatter. The solution is

$$\mathbf{V} = \arg \max_{\mathbf{V}} \frac{|\mathbf{V}^T \mathbf{S}_B \mathbf{V}|}{|\mathbf{V}^T \mathbf{S}_W \mathbf{V}|} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_{H_z}], \quad (8)$$

where $\{\mathbf{v}_{h_z}, h_z = 1, \dots, H_z\}$ is the set of generalized eigenvalues of \mathbf{S}_B and \mathbf{S}_W corresponding to the $H_z (\leq C - 1)$ largest generalized eigenvalues $\{\lambda_{h_z}, h_z = 1, \dots, H_z\}$: $\mathbf{S}_B \mathbf{v}_{h_z} = \lambda_{h_z} \mathbf{S}_W \mathbf{v}_{h_z}$, $h_z = 1, \dots, H_z$. Thus, the LDA feature vector \mathbf{z}_m is obtained as $\mathbf{z}_m = \mathbf{V}^T \mathbf{y}_m$ for the input to a classifier.

The nearest mean classifier (NMC), which assigns label c to the test sample \mathbf{y} if $\bar{\mathbf{y}}_c$ is the class mean nearest to \mathbf{y} , is used with the angle distance measure which is found to significantly outperform the commonly used L1 and L2 measures in gait recognition in [7]. The angle distance between two vectors \mathbf{a} and \mathbf{b} is calculated as $d(\mathbf{a}, \mathbf{b}) = -\frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$. The calculated distances between a sample and the C class means are matched to the interval $[0, 1]$ as required by the AdaBoost.M2 algorithm.

3. EXPERIMENTAL RESULTS

In this section, the proposed solution of boosting LDA on MPCA features (B-LDA-MPCA) is evaluated. In particular, the effects of the gait feature vector dimension H_y and the regularization parameter η are studied, in addition to R , the number of LDA training samples per class. The gait recognition experiments are carried out on the USF HumanID ‘‘Gait Challenge’’ data sets version 1.7 [2]. The human gait sequences in these data sets were captured under different conditions (walking surfaces, shoe types and viewing angles). The gallery set contains 71 sequences (subjects) and seven experiments (probe sets) are designed for human identification. The capturing condition for each probe set is summarized in brackets after the probe name in Table 1, where C,G,A,B,L,R, standing for cement surface, grass surface, shoe type A, shoe type B, left view, and right view, respectively. The capturing condition of the gallery set is GAR. The (third order) tensorial gait samples $\{\mathcal{X}_m\}$ are obtained in the same way as [7] and two examples are shown as unfolded images in Fig. 3. There are 731 gait samples in the Gallery set and each subject has an average of roughly 10 samples available. MPCA is applied to get the gait feature vectors $\{\mathbf{y}_m\}$ for the input to the booster. As in [11], the output dimension H_z of the LDA learner is fixed at 35, which is not optimized, and the maximum number of iterations is set to $T = 60$. The correct classification rate (CCR) is used for recognition performance evaluation.



Fig. 3. Two gait silhouette samples (unfolded).

The best performing set of parameters for the B-LDA-MPCA algorithm is $R = 3$, $H_y = 180$ and $\eta = 10^{-2}$ in our test. The evolutions of various CCRs over the boosting steps are shown in Fig. 4 with this set of parameters. In the

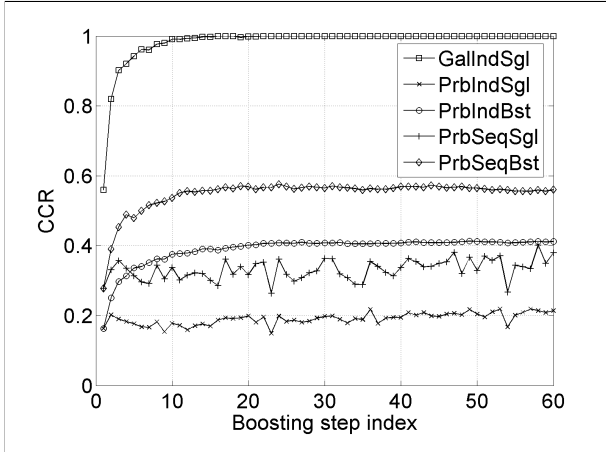


Fig. 4. The evolutions of various CCRs over the boosting steps with the best parameter set.

figure legends, ‘Gal’ means CCRs for the gallery set and ‘Prb’ denotes the average CCRs of the seven probe sets. The CCRs for individual gait samples and gait sequences³ are denoted by ‘Ind’ and ‘Seq’, respectively. The CCRs obtained from the single learner in each step is denoted as ‘Sgl’ and the CCRs obtained from the aggregated learners is denoted by ‘Bst’. For instance, ‘PrbSeqBst’ is the average CCR of all probe sequences obtained from the combined learners. From the figure, the effectiveness of the boosting scheme is observed. The CCRs for the probe samples (sequences) produced by the single learners are around 20% (below 40%), while the CCRs by the boosted learners is around 40% (near 60%), which is a boost of about 20% in the CCR.

Table 1 shows the detailed gait recognition results of B-LDA-MPCA on each probe set and their average with the best parameter set above and for $T = 24$, measured in CCRs on the gait sequences. The results from the baseline algorithm [2] and the best results from the MPCA+LDA algorithm [7] with $H_y = 200$, obtained using the same NMC and angle distance measure, are included in the table for comparison. The best results for each row are highlighted by boldface font in the table. While both the MPCA+LDA and B-LDA-MPCA algorithms outperform the baseline algorithm significantly, the B-LDA-MPCA has improved over the MPCA+LDA consistently on each probe set, with the smallest improvement of 1% on probe A (from 99% to 100%) and the greatest improvement of 5% on probes E and F. The improvement in average CCR is 4%. The consistent improvement shown over both easy probes (A, B, C) and difficult probes (D, E, F, G)⁴ clearly demonstrates the effectiveness of the proposed solution.

³The hypothesis for a sequence is obtained as the average of the hypotheses for its samples.

⁴The probes D, E, F, G are captured on a different surface from the gallery and they are considered to be more challenging sets for the recognition task.

Table 1. Comparison of the (sequence) correct classification rates of the baseline, MPCA+LDA and B-LDA-MPCA algorithms on gait recognition.

Probe	Baseline	MPCA+LDA	B-LDA-MPCA
A (GAL)	79	99	100
B (GBR)	66	85	88
C (GBL)	56	81	85
D (CAR)	29	36	39
E (CBR)	24	29	34
F (CAL)	30	21	26
G (CBL)	10	26	32
Average	42	54	58

The effects of R , H_y and η on the gait recognition performance of the proposed method are shown in Fig. 5. Since it is not possible to show the results of all possible parameter combinations, the effects of a parameter is shown by fixing all the others. None of these parameters is optimized. The fixed set of parameters is chosen to be the best set above: $R = 3$, $H_y = 180$ and $\eta = 10^{-2}$. In the following, only the results ‘PrbSeqBst’ will be shown.

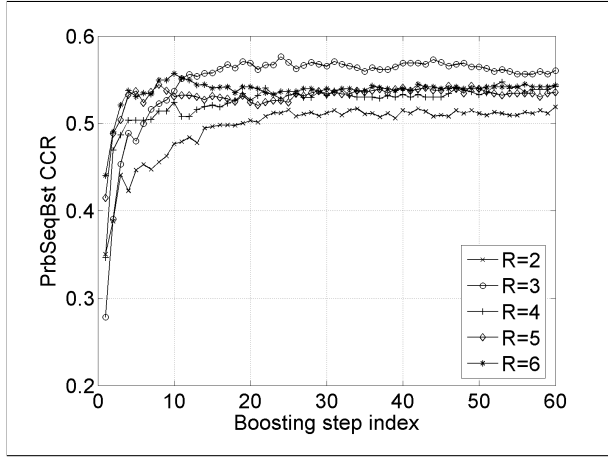
The minimum number of samples in a class is 7 for the gallery set. Therefore, we tested R ranging from 2 to 6 and the results are shown in Fig. 5(a). Similar results as in [11] have been observed here. It shows that the learner cannot be too weak or too strong for the booster to have a positive effect.

This work introduces an additional control of learner weakness by H_y and its effects are shown in Fig. 5(b). As mentioned above, the optimal H_y for a single LDA learner is 200, while the figure shows that the weakened learners with $H_y = 180$ give a much better boosting result than the stronger learners with $H_y = 200$. Hence, H_y affects the booster in a similar way as R .

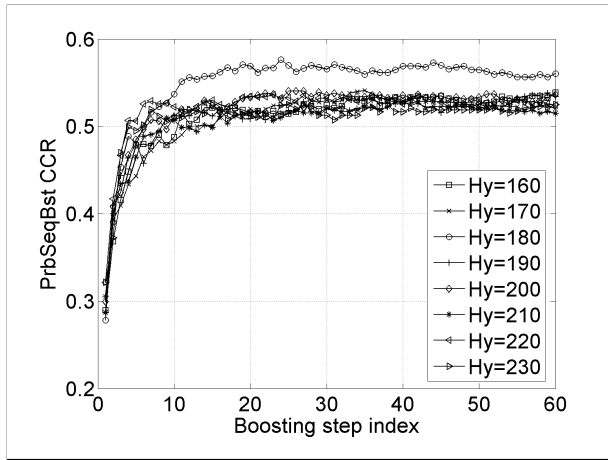
Finally, the effects of the regularization are shown in Fig. 5(c). This figure shows that an appropriate regularization parameter η does result in better generalization. This study confirms our belief that gait recognizer can benefit from making use of the fact that the within-class scatter of gait patterns under various capturing conditions is greater than that under the same capturing condition. It is a promising working direction to train the same booster with different η for recognition in a different capturing condition, with the value η optimized for that particular condition.

4. CONCLUSIONS

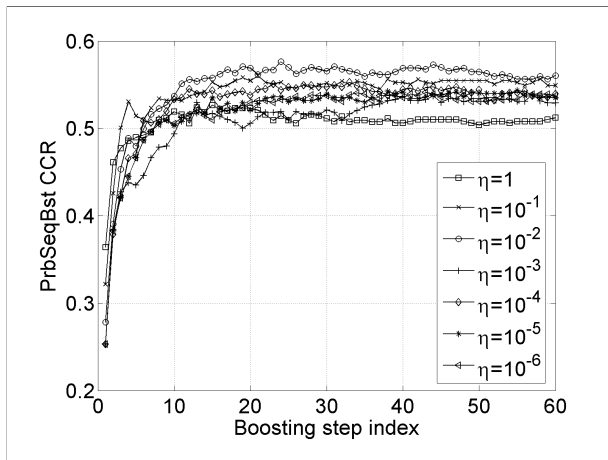
This paper proposes a gait recognition solution through the MPCA [7] and the ensemble-based discriminant learning [11]. The MPCA in [7] is used to extract features from tensorial gait data and a subset of the extracted features are fed into a LDA-style booster. This scheme gives another way of learner weakness control in addition to computational efficiency. The LDA



(a) The effects of R .



(b) The effects of H_y



(c) The effects of η

Fig. 5. The effects of R , H_y and η on the gait recognition performance of the proposed method.

learner in [11] is modified by adopting a simpler weighted pairwise between-class scatter matrix and introducing a regu-

larization term in the within-class scatter matrix so that the gait challenge due to various capturing conditions is taken into account. Experiments carried out on the gait challenge data sets show that the proposed scheme is effective in boosting the gait recognition performance.

5. REFERENCES

- [1] A. Kale, A. N. Rajagopalan, A. Sunderesan, N. Cuntoor, A. Roy-Chowdhury, V. Krueger, and R. Chellappa, "Identification of humans using gait," *IEEE Trans. Image Processing*, vol. 13, no. 9, pp. 1163–1173, Sept. 2004.
- [2] S. Sarkar, P. J. Phillips, Z. Liu, I. Robledo, P. Grother, and K. W. Bowyer, "The human ID gait challenge problem: Data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.
- [3] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, "Gait recognition: a challenging signal processing technology for biometrics," *IEEE Signal Processing Mag.*, vol. 22, no. 6, Nov. 2005.
- [4] C. Y. Yam, M. S. Nixon, and J. N. Carter, "Automated person recognition by walking and running via model-based approaches," *Pattern Recognition*, vol. 37, no. 5, pp. 1057–1072, May 2004.
- [5] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A layered deformable model for gait analysis," in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Apr. 2006, pp. 249–254.
- [6] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Coarse-to-fine pedestrian localization and silhouette extraction for the gait challenge data sets," in *Proc. IEEE Conf. on Multimedia and Expo*, July 2006, pp. 1009–1012.
- [7] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Networks*, vol. 18, no. 6, Nov. 2007, to be published.
- [8] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. the Thirteenth Int. Conf. on Machine Learning*, 1996.
- [9] R. E. Schapire, "The boosting approach to machine learning: An overview," in *Nonlinear Estimation and Classification*, D. D. Denison, M. H. Hansen, C. Holmes, B. Mallick, and B. Yu, Eds. 2003, Springer.
- [10] R. P. W. Duin, M. Skurichina, "Bagging, boosting and the random subspace method for linear classifiers," *Pattern Analysis & Applications*, vol. 5, no. 2, pp. 121–135, 2002.
- [11] J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, and S. Z. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Trans. Neural Networks*, vol. 17, no. 1, pp. 166–178, Jan. 2006.
- [12] M. Loog, R. P. W. Duin, and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise fisher criteria," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 7, pp. 762–766, July 2001.