Structure Sparsity for Multi-camera Gait Recognition

Qiyue Yin[1], Rong Sun[1], Liang Wang[2], and Ran He[2]

[1] Harbin Engineering University
[2] Institute of Automation, Chinese Academy of Sciences
yuejun1001@163.com, sunrong@hrbeu.edu.cn, {wangliang,rhe}@nlpr.ia.ac.cn

Abstract. With the rapid development of surveillance technology, there are often several cameras in one scenario. The multi-camera usage to perform gait recognition becomes a challenge problem. This paper studies multi-camera gait recognition via structure sparsity. For the multi-camera structure in the training set, we propose a structure sparsity algorithm to learn informative and discriminative sparse representations; and for the structure in the testing set, we develop a new classification criteria based on the reconstruction error of learned sparse representations. In addition, we learn a dictionary from the original gait data to further improve recognition accuracy meanwhile reduce computational cost. Experimental results show that the proposed method can efficiently deal with the multi-camera gait recognition problem and outperforms the state-of-the-art sparse representation methods.

Keywords: multi-camera gait recognition, structure sparsity, sparse representation.

1 Introduction

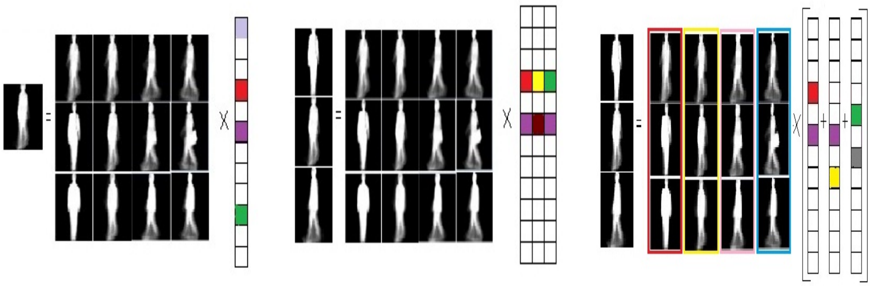
Gait recognition resorts to the walking pattern of humans to perform identity classification [1]. It is the process of classifying an individual at a distance. Since it does not require user cooperation, it favors a growing number of researchers. Generally speaking, gait recognition can be divided into two categories: model-based approaches and appearance-based approaches. Model-based approaches[2][3] apply line, pendulum or 3D model to model the static structure or dynamic movements of human body, and use the model parameters as features. Appearance-based approaches[1]-[6] mainly make use of the silhouette information of human body.

Recently, multi-view gait recognition and compressed sensing based gait recognition have drawn much attention. In [8][9][20], a view transformation model (VTM) was used to transform gait feature from one view (source) into another (target). Zheng et al.[19] applied the low-rank structure of gait to perform robust view transformation. Based on sparse representation classification (SRC) [10][12], Yang et al.[14] extracted averaged boundary by the Canny operator as gait feature, and then performed classification via sparse representation. Experimental results on the USF gait database showed that sparse representation can

C.-L. Liu, C. Zhang, and L. Wang (Eds.): CCPR 2012, CCIS 321, pp. 259–267, 2012.

improve the recognition accuracy. Gong et al.[15] made use of gait energy image (GEI) as gait feature and defined a new distance metric to choose non-polluted area.

With the rapid development of the surveillance technology, there are often several cameras in one scenario. A major challenge in this scenario is that how to use the structure information in both training and testing sets. To the best of our knowledge, there seems few work to discuses this topic. Inspired by L_{21}-norm based group sparsity [16][18], we discuss multi-camera gait recognition via structure sparsity. For the group structure of a training set, we add more prior information to gait recognition and develop an efficient structure sparsity algorithm; and for the prior information of the testing set(views), we propose a new classification scheme based on the reconstruction error. In addition, when using sparse representation for gait recognition, one faces high computation and many zero elements in the gait images. To deal with these two problems, we learn a dictionary for our proposed structure sparsity method, which is proved to be efficient. Fig 1 shows an illustration of our method as compared with other sparse methods.



(a) (b) (c)

Fig. 1. Different frameworks for sparse representation based gait recognition. (a) The SRC framework introduced in [10][12]; (b) the L_{21} framework used in [16][18] that only considers the structure of testing data ; (c) the proposed framework that considers the structural information of both training and testing sets. In the straining set, we assume that gait images belonging to one person are in the same group; and in the testing set, we assume gait images from multiple cameras are in the same group.

Compared with SRC and L_{21}-norm based methods in Fig. 1, the main contributions of our work are two-fold:

1) A novel structure sparsity framework is developed to solve the multi-camera gait recognition problem along with an efficient optimization algorithm.

2) A novel classification criteria is proposed to identify the samples with the same class characteristics through the sum of their reconstruction errors.

The rest of this paper is organized as follows. In section 2, we detail our proposed method. In section 3, the proposed approach is validated by conducting a series of experiments on the well-known CASIA gait dataset, along with the comparison with other sparse techniques. We summarize this paper in Section 4.

2 The Proposed Structure Sparsity Method

In this section, we will give the details of the proposed structure sparsity method. Also a Lagrange multiplier algorithm is used to solve the optimization problem.

2.1 Considering the Structure of the Training Set

We know the training set for the SRC framework is generally redundant, and this means that many samples are of the same characteristics such as the same class label. Once we solve the sparse vector, we can take these samples as a group. Specifically, in the SRC framework, one minimizes the L_1-norm of sparse vector. Here we apply $L_{2,1}$-norm to model the group structure.

Suppose the training set $X = \left[X^1, X^2, \ldots, X^c\right] \in \mathbb{R}^{m*n}$ and $X^i = \left[x_1^i, x_2^i, \ldots, x_{n_i}^i\right] \in \mathbb{R}^{m*n_i}$, where X^i is the set of the i^{th} class; n, n_i and m are the total number of training samples, the total number of the i^{th} class, and the feature dimension respectively. Hence $n = \sum_{i=1}^{c} n_i$. For a test sample y, we solve the following optimization problem:

$$\min_{\beta} \sum\nolimits_{i=1}^{c} \phi(\| \beta^i \|_2) \qquad s.t. \quad y = X * \beta \tag{1}$$

where $\phi(x) = \sqrt{\varepsilon + x^2}$ and β satisfies:

$$\beta = \left[\beta^1, \beta^2, \ldots, \beta^c\right] \in \mathbb{R}^{n*1} \tag{2}$$

where $\beta^i = \left[\beta_1^i, \beta_2^i, \ldots, \beta_{n_i}^i\right] \in \mathbb{R}^{n_i*1}$. We let all the elements of β^i be the same.

In structure sparsity, $\sum_{i=1}^{c} \| \beta^i \|_2$ is also called $L_{2,1}$-norm [16][18] which can result in less nonzero elements entering the class and thus suppressing their contributions to the reconstruction error.

2.2 Considering the Structure of the Testing Set

For the testing data with similar characteristics such as the same class label, we consider all of input testing samples and present a novel classification criteria instead of solving the sparse matrix. A simple idea is to sum the reconstruction error of samples which has the same characteristics such as same class label. Then we will classify all these test samples to the class which have the minimum sum of reconstruction errors. The method is formulated as follows.

For the test samples with the same characteristics such as the same class label$Y^i \in \{y_1^i, y_2^i, \ldots, y_{k_i}^i\}$, where k_i is the total number of test samples of the i^{th} class. We calculate the reconstruction error of each test sample:

$$r_i(y_k^i) = \| y_k^i - X_i * \beta_k^i(i) \|_2 \tag{3}$$

where $\beta_k^i(i)$ is the sparse coding vector associated with y_k^i. The sum of the reconstruction errors of class i is:

$$r_i(Y^i) = \sum\nolimits_{k=1}^{k_i} r_i(y_k^i) \tag{4}$$

And we will give the class label of the samples based on:

$$\text{identify}(Y^i) = \arg\ \min_i r_i(Y^i) \tag{5}$$

It is worth emphasizing that we can give a low weight when summing the reconstruction error when the testing data is corrupted, thus increasing the credibility of gait recognition .

2.3 An Efficient Algorithm for the Proposed Structure Sparsity

Since the proposed structure sparsity problem in (1) is difficult to be directly optimized, we develop an efficient algorithm to solve it. Fortunately, we can resort to the half-quadratic optimization in [18] by introducing a relaxation ε. According to [18][17], we have the following Lemma.

Lemma 1. *Let $\phi(.) = \sqrt{\varepsilon + \|\beta^i\|_2^2}$, there exists a dual potential function $\varphi(.)$, such that*

$$\phi(\|\beta^i\|_2) = \min_p \left\{p \|\beta^i\|_2^2 + \varphi(p)\right\}$$

where p is determined by a minimizer function with respect to $\phi(.)$.

Based on Lemma 1, we have the augmented objective of (1) with respect to β and p,

$$\min_{\beta,p} \sum\nolimits_{i=1}^{c} (p_i \| \beta^i \|_2^2 + \varphi(p_i)) \qquad s.t. \quad y = X * \beta \tag{6}$$

Then we can alternately minimize the above problem as,

$$\begin{gathered} p^t = \delta(\| \beta^i \|_2) \\ \beta^t = \arg\min_\beta \sum\nolimits_{i=1}^{c} p_i^t \| \beta^i \|_2^2 \qquad s.t. \quad y = X * \beta \end{gathered}$$

where the minimizer function $\delta(x) = 1/\sqrt{\varepsilon + x^2}$ [18].

By using the Lagrange multiplier method, we can simplify the second problem by alternately solving,

$$L(\beta) = \beta^T P \beta - \lambda(X\beta - y) \tag{7}$$

where P is a diagonal matrix with the j^{th} diagonal element $P_{jj} = 1/\sqrt{\varepsilon + \|\beta^i\|_2^2}$ (x_j belongs to the i-th class). Taking the derivative of β and setting the derivative to zero, we obtain:

$$\partial L(\beta)/\partial(\beta) = P\beta - X^T\lambda = 0 \tag{8}$$

Left multiplying the two sides of Eq.(7) by XP^{-1} and using the equality contraint $X\beta = y$, we have:

$$\lambda^T = (XP^{-1}X^T)^{-1} * y \tag{9}$$

Substitute Eq. (8) into Eq. (7), we have:

$$\beta = P^{-1}X^T(XP^{-1}X^T)^{-1}y \tag{10}$$

Now the β is the sparse vector β_k^i corresponding to the testing sample y_k^i.

We can use the following iterative algorithm to solve the final β, and it is noteworthy that the elements of β satisfy the condition in Section 2.1.

Data **:** $A \in \mathbb{R}^{m*n}$, $y \in \mathbb{R}^{m*1}$
Result **:** $\beta \in \mathbb{R}^{n*1}$
Set $t = 0$, initialize P as an identity matrix
Repeat
Calculate $\beta = P^{-1}A^T(AP^{-1}A^T)y$.
Calculate the diagonal matrix P, $P_{jj} = \frac{1}{\|\beta^i\|_2}$, $\forall \sum_{i=1}^{i-1}(n_i) < j \leq \sum_{i=1}^{i}(n_i)$.
$t = t + 1$.
Until converges.

2.4 A Learned Dictionary for the Proposed Structure Sparsity

When using the proposed structure sparsity for gait recognition, it takes a high computational cost because of the high dimensions of gait images. And this becomes harder when the gait matrix is irreversible. So it is essential for us to learn an efficient dictionary for gait recognition.

The easiest way to remove the nonessential elements is linear dimensionality reduction [7][13][17]. We will learn a PCA dictionary which use PCA to reduce the dimensions of the training data and testing data. The principal components of PCA are widely used as eigenface, and it means the principal components can reconstruct the human face well. The proposed structure sparsity is also a process of reconstruction for the testing data,thus the PCA method is suitable for our problem. In contrast, we also give the usually used LDA dictionary and $L_{21} - norm$ dictionary, where the $L_{21} - norm$ introduced in[16] is used to learn a subspace for comparison.

3 Experiments

In this section, we use the proposed structure sparsity for multi-camera gait recognition on the CASIA gait database. For the dataset, there are 124 individuals walking in three ways, .i.e., walking normally(nm), walking with a bag(bg) and walking with a coat(cl). And there are up to 6 nm sequences, 2 bg sequences and 2 cl sequences with eleven viewpoints of each person. Each sequence contains some cycle of gait images, and we will use the GEI template[4] to do experiments because it saves both storage space and computation time for recognition and is less sensitive to silhouette noise in individual frames.

3.1 The Proposed Structure Sparsity for Multi-view Gait Recognition

In the experiments, we randomly choose 100 subjects as the training set with each subject containing 4 sequence of nm images with 11 viewpoints, thus we have 4400 samples for the training set. For the testing set we choose the other six sequences with 11 viewpoints, .i.e., two nm sequences, two bg sequences and two cl sequences of the same subjects.

The main reason of such experimental setting is due to the multi-camera gait recognition problem. Besides, in real applications, we only collect videos that the trainers walk normally without wearing coat or bag. Although it greatly increases the difficulty of identification when we set bg and cl as the testing set, they are the case we often face. It is also worth mentioning that we often get many views of the same person through different cameras in practical application. Thus the assumption that we take all the 11 viewpoints of a person as a group is reasonable, or at least we take few views of a person as a group. And for the training data, every person has 11 viewpoints, and each viewpoint has 4 nm GEI images. We take the 4 GEI images of the same viewpoint as a structure. Using the PCA dictionary, we have the following experimental results.

We compare our method with the other two methods which use the sparse theory on multi-camera gait recognition. And the nearest neighbor (NN) algorithm is used as baseline. The former two methods are SRC introduced by John Wright[10] and group sparsity proposed in [16] for classification. The results are the average classification accuracy of the sequences of nm, bg and cl.

Table 1. The proposed structure sparsity for multi-view gait recognition

Test data	NN	SRC	group sparsity	proposed
walking normally(nm)	99.68%	99.78%	**100.00%**	**100.00%**
walking with a bag(bg)	49.82%	62.82%	81.00%	**88.50%**
walking with a coat(cl)	17.05%	37.46%	43.00%	**59.50%**

Table 1 lists the comparison among four methods. For the group sparsity method, we still use the reconstruction error for the classification of the testing samples and it considers the structure of testing data only. We can see the proposed structure sparsity method gains the highest recognition rate for multi-view gait recognition. It is no wonder the proposed method outperforms the other ones because of more prior information considered.

In table 2, 'proposed1' represents the proposed structure sparsity which considers the structure of the training set only and 'proposed2' with the consideration of testing data only. For the 'proposed1' method, we add prior knowledge of training data, and thus it performs better for granted than SRC overall. For the comparison of the proposed2 to group sparsity, maybe the classification criteria is more suitable here for multi-view gait recognition. It is worth emphasizing that the proposed method again obtains the best result for multi-view gait recognition here.

Table 2. A comparison by taking the structure sparsity into two parts

Test data	SRC	proposed1	group sparsity	proposed2	proposed
walking normally(nm)	**99.78%**	99.68%	**100.00%**	**100.00%**	**100.00%**
walking with a bag(bg)	62.82%	**63.14%**	81.00%	**85.50%**	**88.50%**
walking with a coat(cl)	37.46%	**42.96%**	43.00%	**55.00%**	**59.50%**

3.2 A Comparison of the Dictionary Learning

Here we compare the PCA method and the LDA method, which are widely used in sparse representation. The dimension of PCA is 200(the original is 3600) and it is selected based on the accumulation of the eigenvalue of $cov(X)$. A dimension of 200 can highly reduce the computational complexity and remove noise of the GEI images. And one more method in[16] is used for dictionary learning. In the work of [16], the proposed method for feature selection is robust. The three dictionary learning methods are tested using both the baseline NN algorithm and SRC algorithm.

Table 3. A comparison of the dictionary learning on NN and SRC algorithm

Test data	PCA+NN	LDA+NN	L_{21}+NN	PCA+SRC	LDA+SRC	L_{21}+SRC
nm	99.68%	**99.73%**	98.00%	**99.78%**	**99.78%**	99.56%
bg	49.82%	55.39%	**65.23%**	**62.82%**	51.91%	52.96%
cl	17.05%	22.68%	**45.41%**	**37.46%**	20.50%	35.32%

From the left of Table 3, we can observe that PCA+NN is worse than the other two methods. When the nearest neighbor classifier is used, a supervised learning method may be better than a unsupervised one. This is because supervised methods make use of discriminative information to learn a subspace. However, it is interesting to observe from the right of Table 3 that PCA+SRC always achieves the highest recognition rates, just as the analysis we explain in subsection 2.4. And here the testing data conclude bg sequences and cl sequences which mean a big damage for the training data, so the supervised methods may not perform well as they used to be. In summary, the PCA method is suitable for sparse representation based gait recognition.

4 Conclusion

In this work, we have developed a structure sparsity method for multi-camera gait recognition, which combines prior information from both testing data and training data. For the training data, we assumed that training datum with same characteristics belong to a group. And for the testing data, we proposed a new classification criteria in which all the test samples with the same characteristics are assumed to be the class corresponding to the minimum sum of reconstruction errors. We also learned an effective dictionary for the proposed structure sparsity method. Experimental results show that our method can efficiently deal with the multi-camera gait recognition problem.

Acknowledgment. This work is jointly supported by National Natural Science Foundation of China (61175003, 61103155), Hundred Talents Program of CAS, the strategic Priority Research Program of CAS (XDA06030300), and the National Basic Research Program of China (2012CB316300).

References

1. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette Analysis-based Gait Recognition for Human Identification. TPAMI 25(12), 1505–1518 (2003)
2. Zhang, R., Vogler, C., Metaxas, D.: Human Gait Recognition. In: CVPRW, 18 pages (2004)
3. Zhao, G., Liu, G., Li, H., Pietikainen, M.: 3D Gait Recognition using Multiple Cameras. In: FGR, pp. 529–534 (2006)
4. Han, J., Bhanu, B.: Individual Recognition Using Gait Energy Image. TPAMI 28(2), 316–322 (2006)
5. Tao, D., Li, X., Wu, X., Maybank, S.J.: General Tensor Discriminant Analysis and Gabor Features for Gait Recognition. TPAMI 29(10), 1700–1715 (2007)
6. Wang, C., Zhang, J., Wang, L., Pu, J., Yuan, X.: Human Identification Using Temporal Information Preserving Gait Template. TPAMI (99) (2011)
7. Lei, Z., Chu, R.F., He, R., Liao, S.C., Li, S.: Face Recognition by Discriminant Analysis with Gabor Tensor Representation. In: Advances in Biometrics (2007)
8. Kusakunniran, W., Wu, Q., Zhang, J., Li, H.: Support Vector Regression for Multi-view Gait Recognition based on Local Motion Feature Selection. In: CVPR, pp. 974–981 (2010)
9. Kusakunniran, W., Wu, Q., Zhang, J., Li, H.: Gait Recognition under Various Viewing Angles based on Correlated Motion Regression. TCSVT (99), 1–1 (2012)
10. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust Face Recognition via Sparse Representation. TPAMI 31(2), 210–227 (2009)
11. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T.S., Yan, S.: Sparse Representation for Computer Vision and Pattern Recognition. Proceedings of the IEEE 98(6), 1031–1044 (2010)
12. He, R., Zheng, W.-S., Hu, B.-G., Kong, X.W.: A Regularized Correntropy Framework for Robust Pattern Recognition. Neural Computation 23(8), 2074–2100 (2011)
13. He, R., Hu, B.-G., Zheng, W.-S., Kong, X.W.: Robust Principal Component Analysis Based on Maximum Correntropy Criterion. IEEE Transactions on Image Processing 20(6), 1485–1494 (2011)
14. Yang, Q., Xue, D.-Y., Cui, J.-J.: Gait Recognition Using Sparse Representation. Journal of Northeastern University(Natural Science), 136–139 (2012)
15. Gong, M., Xu, Y., Yang, X., Zhang, W.: Gait Identification by Sparse Representation. FSKD 3, 1719–1723 (2011)

16. Nie, F., Huang, H., Cai, X., Ding, C.: Efficient and Robust Feature Selection via Joint $l_{2,1} - Norms$ Minimization. In: NIPS, vol. 23, pp. 1813–1821 (2010)
17. He, R., Sun, Z.N., Tan, T.N., Zheng, W.S.: Recovery of Corrupted Low-Rank Matrices via Half-Quadratic based Nonconvex Minimization. In: IEEE CVPR (2011)
18. He, R., Tan, T.N., Wang, L., Zheng, W.S.: L21 Regularized Correntropy for Robust Feature Selection. In: IEEE CVPR (2012)
19. Zheng, S., Huang, K.Q., Tan, T.N., He, R., Zhang, J.: Robust View Transformation Model for Gait Recognition. In: ICIP (2011)
20. Kusakunniran, W., Wu, Q., Li, H.D., He, R., Zhang, J.: Multiple Views Gait Recognition Using View Transformation Model Based on Optimized Gait Energy Image. In: IEEE ICCV (2009)