Multi-View Pedestrian Recognition using Shared Dictionary Learning with Group Sparsity

Shuai Zheng[1], Bo Xie[2], Kaiqi Huang[1], and Dacheng Tao[3]

1. National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing, China
`szheng@nlpr.ia.ac.cn,kqhuang@nlpr.ia.ac.cn`
2. Chinese University of Hong Kong, Hongkong, China
`zixu1986@gmail.com`
3. Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia
`dacheng.tao@uts.edu.au`
Office number: +61 2 9514 1829, Fax number: +61 2 9514 4517

Abstract. Pedestrian tracking in multi-camera is an important task in intelligent visual surveillance system, but it suffers from the problem of large appearance variations of the same person under different cameras. Inspired by the success of existing view transformation model in multi-view gait recognition, we present a novel view transformation model based approach named shared dictionary learning with group sparsity to address the problem. It projects the pedestrian appearance feature descriptor in probe view into the gallery one before feature descriptors matching. In this case, $L_{1,\infty}$ regularization over the latent embedding ensure the lower reconstruction error and more stable feature descriptors generation, comparing with the existing Singular Value Decomposition. Although the overall optimization function is not global convex, the Nesterovs optimal gradient scheme ensure the efficiency and reliability. Experiments on VIPeR dataset show that our approach reaches the state-of-the-art performance.

Key words: multiview learning, dimension reduction, stochastic neighbor embedding, image retrieval

1 Introduction

Recognizing humans at a distance is not only a popular topic in computer vision community, but also very important in intelligent visual surveillance applications. However, this task is very challenging since there are tremendous variations in appearance and structure under different illumination, poses, occlusions and viewpoints [1–3]. The complex intra-class variations are difficult to model and lead to recognition failures. Besides, in uncontrolled scenarios, people may carry bags or exhibit other unexpected noises, which further deteriorate performance.

In this paper, we study human recognition in the multi-view setting. It refers to identifying a person across different viewpoints. This problem arises in numerous computer vision scenarios, such as pedestrian re-identification, multi-camera tracking, etc.
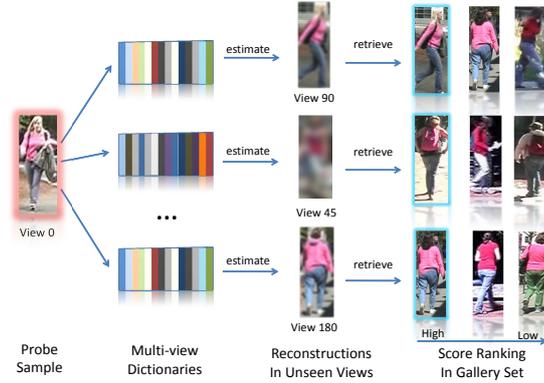
Fig. 1. Illustration of the multi-view human recognition system. Learned SDL-GS model is used to estimate data in unseen views given probe samples. In each unobserved view, the Nearest Neighbor (NN) rule is used as a classifier to retrieve samples in the gallery set.

The major issue is how to learn a view invariant representation while preserving discriminative information for accurate recognition. One paradigm is to use features such as Self Similarity Plots (SSP) [4] and flow fields gait representation [5]. Color histograms and texture filters are used in [6, 7, 5] to recognize pedestrian across different views. However, much discriminative information is lost in the process to achieve viewpoint invariance. For controlled scenes, viewpoint differences can be handled by view synthesis using planar imaging geometry or self calibration [8]. But this requires explicit knowledge of the scenes, which is unknown in most practical cases. Dikmen [7] proposed to learn a suitable metric where variations caused by view changes are offset by supervised learning. While label information is very important for recognition [5], neglecting the relationships between views is prone to sub-optimal solutions since conflicting views may be forced together simply because they are labeled as similar. Moreover, View Transformation Models (VTM) based on linear factorization [9, 10] and Gaussian processes [8] also demonstrate satisfactory performance in multi-view gait recognition. However, these techniques are sensitive to noise and are likely to overfit in cases of small training sets.

When inferring the sparse latent embedding, traditional sparse coding techniques suffer from random locations of non-zero coefficients. We propose to use $l_{1,\infty}$ regularization on the latent embedding for group sparsity effect. Instead of being randomly distributed, the locations of non-zero coefficients will tend to cluster into groups. Such effect is resistant to noise disturbance and leads to improved recognition performance. Moreover, learning and inference in our approach only consists of convex optimization, which can be efficiently solved by Nesterov's optimal gradient scheme. Fig. 1 illustrates the mechanism of our approach at recognition time.

The contributions of our work are three-fold: 1) SDL-GS is capable of accurately estimating test samples in unseen views, so as to handle large variations; 2) Group sparsity leads to stable latent embedding and increases recognition performance significantly; 3) The computation of SDL-GS only consists of convex optimization and can be carried out efficiently with global optimum guarantee.

The remainder of the paper is organized as follows. In Section 2, we present our multi-view dictionary learning formulations. Group sparsity with $l_{1,\infty}$ norm regularization is introduced in Section 3. In Section 4, we evaluate our approach on VIPeR [6]. Finally, we conclude the paper in Section 5.

2 Shared Dictionary Learning with Group Sparsity

In this section, we first briefly review sparse coding techniques. Then we introduce our dictionary learning algorithm for problems with multiple views.

2.1 Sparse Coding

Recently, sparse coding or sparse representation is very popular within the face and object recognition communities[11]. It is reported that the reconstruction error of sparse coding is much lower than that of Singular Value Decomposition (SVD). Besides, sparse coefficients can be less sensitive to noise and occlusion [12]. But sparse coding has its drawback that the positions of non-zero coefficients in descriptor vectors are not stable, which leads it may not be robust enough.

Specifically, let $\mathbf{x}_i \in \mathbb{R}^d$ denote the i-th input data, which can be vectorized image pixels or extracted features. And denote $\mathbf{D} \in \mathbb{R}^{d \times k}$ as the dictionary matrix. We seek a sparse latent embedding to represent $\mathbf{x}_i$,

$$\boldsymbol{\alpha}_i = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}\|^2 + \lambda\|\boldsymbol{\alpha}\|_1 \tag{1}$$

where $\|\boldsymbol{\alpha}\|_1 = \sum_{j=1}^{k} |\alpha_j|$ is the l_1 norm and λ is the coefficient that balances the trade-off between the reconstruction error term and the regularization term.

2.2 Multi-view Dictionary Learning with Shared Latent Embedding

In our multi-view dictionary learning approach, we assume the sparse latent embedding is shared across all views, and each view-specific dictionary is responsible for generating the corresponding data samples.

This generative model provides a straightforward way of estimating unseen views given a subset of observed views. And our sparse dictionary-based approach can significantly reduce reconstruction error on unseen views and improve recognition accuracy compared with simple factorization approaches such as Singular Value Decomposition [9].

In details, given N data, each with V views, the goal is to learn V view-specific dictionaries that fit the data well. Let $\mathbf{x}_i^v$ denote the i-th data in the v-th view and $\mathbf{D}^v$ denote the dictionary for the v-th view. The dictionary learning approach is formulated as the following problem

$$\min_{\mathbf{D}^v, \boldsymbol{\alpha}_i} \frac{1}{V} \sum_{v=1}^{V} \sum_{i=1}^{N} \|\mathbf{x}_i^v - \mathbf{D}^v \boldsymbol{\alpha}_i\|^2 + \lambda \sum_{i=1}^{N} \|\boldsymbol{\alpha}_i\|_1 \tag{2}$$
$$s.t. \|\mathbf{d}_j^v\| \leq 1, \ 1 \leq j \leq k$$

where $\mathbf{d}_j^v$ is the j-th column of $\mathbf{D}^v$. Note that $\boldsymbol{\alpha}_i$ is shared across all views.

We denote the approach characterized in Eq. 3 as Shared Dictionary Learning with Sparse Coding (SDL-SC). SDL-SC is solved by alternating optimization. Fixing $\boldsymbol{\alpha}_i$'s, solving $\mathbf{D}^v$ is a quadratic programming problem and can be efficiently solved by Block Coordinate Descent [13]. Given $\mathbf{D}^v$, solving $\boldsymbol{\alpha}_i$'s is traditional sparse coding and can be efficiently carried out by approaches like LARS/Lasso [14] or feature-sign [15].

2.3   Inference on the Latent Embedding

Once we have learnt the set of dictionaries $\{\mathbf{D}^v\}_{v=1}^V$, we infer the shared latent embedding for a new test sample $\hat{\mathbf{x}}_i^v$ by solving the following problem

$$\hat{\boldsymbol{\alpha}}_i = \arg\min_{\hat{\boldsymbol{\alpha}}} \frac{1}{V} \sum_{v=1}^V \|\hat{\mathbf{x}}_i^v - \mathbf{D}^v \hat{\boldsymbol{\alpha}}\|^2 + \lambda\|\hat{\boldsymbol{\alpha}}\|_1 \qquad (3)$$

In some scenarios, only a subset of views $\mathcal{V}$ are available and we would like to reconstruct other views from the given data. This can be done by first inferring the latent embedding $\tilde{\boldsymbol{\alpha}}_i$ from the given views $\hat{\mathbf{x}}_i^v$, $v \in \mathcal{V}$

$$\min_{\tilde{\boldsymbol{\alpha}}_i} \frac{1}{|\mathcal{V}|} \sum_{v\in\mathcal{V}} \|\hat{\mathbf{x}}_i^v - \mathbf{D}^v \tilde{\boldsymbol{\alpha}}_i\|^2 + \lambda\|\tilde{\boldsymbol{\alpha}}_i\|_1 \qquad (4)$$

where $|\mathcal{V}|$ is the cardinality of the set $\mathcal{V}$.

Then data in the unseen views can be estimated through $\tilde{\mathbf{x}}_i^v = \mathbf{D}^v \tilde{\boldsymbol{\alpha}}_i$, where $v \notin \mathcal{V}$.

3   Group Sparsity Induced by $l_{1,\infty}$ Norm

One problem with sparse coding is that the locations of non-zero coefficients of latent embedding are sensitive to noise – two similar data samples may have quite different sparse coding coefficients, thus making the sparse coding scheme unstable for similarity measure. Recently, Huang [16] proposed a dynamic group sparsity algorithm that can incorporate the prior of clustering trend of non-zero elements. Their approach is a greedy sparse recovery algorithm, with iterative pruning.

We introduce a group sparsity formulation with $l_{1,\infty}$ norm regularization. Compared with the pruning approach, our convex alternative is much simpler to implement and can take advantage of any existing convex optimization solvers. More importantly, our approach does not require any knowledge about the locations or sizes of the group clusters and can achieve superior results than traditional sparse coding techniques in almost all tasks.

Our formulation is similar to group sparse coding [17]. However, our approach adopts $l_{1,\infty}$ norm regularization, which is reported to exhibit better performance than $l_{1,2}$ norm [18]. Jia [19] also proposed a factorized multi-modal learning framework that places group sparsity constraints on the dictionaries. While these constraints are sensible on multi-modal data where data from different modalities are drastically different, we have observed better performance by simple norm constraints for multi-view data where different views have much in common.

3.1 Convex Group Sparsity

In order to faciliate the use of $l_{1,\infty}$ norm, we first represent data in matrix form. Let $\mathbf{X}^v = [\mathbf{x}_1^v, \mathbf{x}_2^v, \cdots, \mathbf{x}_n^v]$ and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n]$. The dictionary learning problem is thus reformulated as

$$\min_{\mathbf{D}^v, \boldsymbol{\alpha}} H(\mathbf{D}^v, \boldsymbol{\alpha}) = \frac{1}{NV} \sum_{v=1}^V \|\mathbf{X}^v - \mathbf{D}^v \boldsymbol{\alpha}\|_{\text{Fro}}^2 + \lambda \|\boldsymbol{\alpha}\|_{1,\infty} \tag{5}$$

$$s.t. \ \|\mathbf{d}_j^v\| \le 1, \ 1 \le j \le k$$

where $\|\cdot\|_{\text{Fro}}$ is the Frobenius norm. The $l_{1,\infty}$ norm is defined as $\|\boldsymbol{\alpha}\|_{1,\infty} = \sum_{i=1}^k \max_j |\alpha_i^j|$, where α_i^j is the (i,j)-th element in $\boldsymbol{\alpha}$.

The property of $l_{1,\infty}$ norm naturally leads to entire rows of $\boldsymbol{\alpha}$ to be zeroed out. Therefore, the locations of non-zero elements will cluster into groups instead of being randomly distributed. Such grouping effect is observed to bring significant performance boost compared with traditional sparse coding scheme.

Algorithm 1 Multi-view Shared Dictionary Learning

Input: $\mathcal{X} = \{\mathbf{X}^1, \cdots, \mathbf{X}^V\}$.
Output: $\mathcal{D} = \{\mathbf{D}^1, \cdots, \mathbf{D}^V\}$.
Initialize: $\boldsymbol{\alpha}_0 \in \mathbb{R}^{k \times N}, \mathbf{X}_{cat} = [\mathbf{X}^1; \cdots; \mathbf{X}^V]$.

Iterate for $t = 0, 1, 2, \ldots$ until convergence:

1. For $v \leftarrow 1$ to V do

$$\mathbf{D}_t^v = \arg\min_{\mathbf{D}^v} \|\mathbf{X}^v - \mathbf{D}^v \boldsymbol{\alpha}_t\|_{\text{Fro}}^2, \ s.t. \ \|\mathbf{d}_j^v\| \le 1$$

   End for

2. Concatenate dictionaries vertically

$$(\mathbf{D}_{cat})_t = \left[\mathbf{D}_t^1; \cdots; \mathbf{D}_t^V\right]$$

3. Compute $\boldsymbol{\alpha}_{t+1}$ by using Algorithm 2

$$\boldsymbol{\alpha}_{t+1} = \arg\min_{\boldsymbol{\alpha}} \frac{1}{NV} \left\|\mathbf{X}_{cat} - (\mathbf{D}_{cat})_t \boldsymbol{\alpha}\right\|_{\text{Fro}}^2 + \lambda \|\boldsymbol{\alpha}\|_{1,\infty}$$

The optimization of (SDL-GS) is solved by alternating optimization. At every iteration, we 1) solve each view-specific dictionary $\mathbf{D}^v$ with fixed $\boldsymbol{\alpha}$ by using Block Coordinate Descent [13] and then 2) compute $\boldsymbol{\alpha}$ by using Nesterov's optimal gradient scheme as in Algorithm 2.

We adopt a window based stopping criterion: for a given window size h, at every iteration t, we calculate the following ratio

$$r = \frac{\max \mathcal{W}_t - \min \mathcal{W}_t}{\max \mathcal{W}_t}$$

where the set $\mathcal{W}_t = \{H_{t-h+1}, \cdots, H_t\}$ consists of history objective values in a window. If $r < \theta$, where θ is a predefined value, the algorithm stops iterating.

We illustrate the detailed procedures in Algorithm 1. SDL-SC as in Eq. 3 is solved similarly, but with l_1 norm over $\boldsymbol{\alpha}$, thus more sensitive to noise.

3.2 Nesterov's Optimal Gradient Scheme for $l_{1,\infty}$ Regularization

Now, we present an accelerated gradient descent methhod [20] to solve the convex $l_{1,\infty}$ regularization problem

$$\min_{\boldsymbol{\alpha}} F(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}) + \Psi(\boldsymbol{\alpha})$$

$$= \frac{1}{NV} \sum_{v=1}^{V} \|\mathbf{X}^v - \mathbf{D}^v \boldsymbol{\alpha}\|_{\mathrm{Fro}}^2 + \lambda \|\boldsymbol{\alpha}\|_{1,\infty} \tag{6}$$

In order to apply Nesterov's optimal gradient scheme, we first need to define the proximal function and the generalized gradient mapping

$$Q_L(\boldsymbol{\alpha}, \boldsymbol{\alpha}_k) = f(\boldsymbol{\alpha}_k) + \langle \nabla f(\boldsymbol{\alpha}_k), \boldsymbol{\alpha} - \boldsymbol{\alpha}_k \rangle$$

$$+ \frac{L}{2} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\|_{\mathrm{Fro}}^2 + \lambda \|\boldsymbol{\alpha}\|_{1,\infty} \tag{7}$$

$$q_L(\boldsymbol{\alpha}_k) = \arg\min_{\boldsymbol{\alpha}} Q_L(\boldsymbol{\alpha}, \boldsymbol{\alpha}_k) \tag{8}$$

where L is the a positive constant.

The proposed algorithm then alternately updates 1) a latent embedding matrix sequence $\{\boldsymbol{\alpha}_k\}$ by generalized gradient mapping and 2) an aggregation matrix sequence $\{\boldsymbol{\beta}_k\}$ by linear combination of $\{\boldsymbol{\alpha}_k\}$

$$\boldsymbol{\alpha}_{k+1} = q_{L_{k+1}}(\boldsymbol{\beta}_k) \tag{9}$$

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\alpha}_{k+1} + \frac{c_k - 1}{c_{k+1}} (\boldsymbol{\alpha}_{k+1} - \boldsymbol{\alpha}_k) \tag{10}$$

where $\{c_k\}$ is a scalar sequence and L_k is adjusted in every iteration.

The detailed description of Nesterov's optimal gradient is illustrated in Algorithm 2. The stop criterion is the same as used in Algorithm 1. Computing $q_L(\boldsymbol{\beta}_k)$ in Algorithm 2 can efficientlybe done by projection onto the l_∞ ball. [18].

4   Experiment

In this section, we demonstrate the effectiveness of the proposed SDL-GS algorithm on a VIPeR dataset[6], comparing with the state-of-the-art approaches. VIPeR dataset is most popular Multi-view camera tracking public available dataset. This dataset contains 632 unique pedestrians and a total of 1264 images with two views per pedestrian.

Algorithm 2 Nesterov's Optimal Gradient Scheme

Input: λ, L_0, η, $\mathbf{X}_{cat} \in \mathbb{R}^{D \times N}$, $\mathbf{D}_{cat} \in \mathbb{R}^{D \times k}$, where D is the concatenated feature dimension as in Algorithm 1.
Output: $\boldsymbol{\alpha} \in \mathbb{R}^{k \times N}$.
Initialize: $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0 = \boldsymbol{\alpha}_0$, $c_0 = 1$.

Iterate for $k = 0, 1, 2, \ldots$ until convergence of $\boldsymbol{\alpha}_k$:

1. Set $\bar{L} = L_k$
2. While $F\left(q_{\bar{L}}\left(\boldsymbol{\beta}_k\right)\right) > Q\left(q_{\bar{L}}\left(\boldsymbol{\beta}_k\right), \boldsymbol{\beta}_k\right)$

$$\bar{L} = \eta \bar{L}$$

   End while
3. Set $L_{k+1} = \bar{L}$ and update

$$\boldsymbol{\alpha}_{k+1} = q_{L_{k+1}}(\boldsymbol{\beta}_k)$$
$$c_{k+1} = \frac{1 + \sqrt{1 + 4c_k^2}}{2}$$
$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\alpha}_{k+1} + \frac{c_k - 1}{c_{k+1}}\left(\boldsymbol{\alpha}_{k+1} - \boldsymbol{\alpha}_k\right)$$

Implementation Details and Evaluation Experiments on this dataset are carried out in two aspects: 1) to compare SDL-GS with other factorization models, such as SVD and SDL-SC; 2) to compare our approach with other paradigms, such as AdaBoost and metric learning. All the algorithms are evaluated in terms of the Cumulative Matching Characteristic (CMC) curve, the same as in [6].

The images in the VIPeR dataset are 128 pixels high and 48 pixels wide. We extract color histograms from 8×8 pixels blocks. The blockes are densely collected from a regular grid with 4-pixel spacing in the vertical direction and 8-pixel spacing in the horizontal direction. The blocks are overlapped, with a step size of 4 pixels. We use RGB and HSV color spaces and extract 8-bin histograms in each channel. The all the histograms over the grid are concatenated to represent an image. The feature vector size is 2232 for both RGB and HSV. We also adopt other shape and texture feature descriptors, including dense SIFT descriptors [21] and dense Local binary pattern (LBP) descriptors[22]. The blocks used in these features is 8×8 pixels with 4 pixels spacing. We concatenate all the feature descriptors together and then apply Principal Component Analysis (PCA) to obtain 60-dimension low-dimension representation. The reduced features are normalized to unit vectors.

Since our view transformation model need to be trained on each combination of probe view and gallery view, we separate datasets into training set and testing set in each subsets. We learn each VTM respected to the different combinations of probe and gallery views, and calculate the cumulative recognition rates based on matching results from each view transformation model. After transform, we conduct the template matching between transformed data and gallery data to obtain the matching scores. The people number distribution of different combinations of probe viewpoints and gallery

viewpoints are not balanced. We directly employ template matching on those combinations that contains only 1, 3 or 4 pairs data (e.g. probe view is 90 degree while gallery view is 0 degree, probe 45 and gallery 45).Since there are some result variations in spliting training sets and testing sets, we report the average performanec of the approaches

Comparison within Factorization Models To validate the effectiveness of proposed approach, we compare with the other factorization approaches, SVD and SDL-SC, on each subsets of VIPeR dataset. In each subset, the probe data all come from the same viewpoints while gallery data are collected in the same way. We select the half of the data to constitute the training set while half of the data to constitute the testing set. Recognition rates at different retrieving ranks are presented in Table 1. And SDL-GS is superior than sparse coding but some times it is comparable with SVD.

Table 1. Recognition performance comparison among SDL-GS, SDL-SC and SVD in different subsets of the VIPeR dataset. The parameters of SDL-GS and SDL-SC are tuned to optimal performance. In the Viewpoint Pairs column, $0°$ to $90°$ refers to a subset of data where probe samples are in the viewpoint angle $0°$ and gallery samples are in the viewpoint angle $90°$. The other two notations follow the same convention. "–" denotes the "not applicable" because the subset does not have enough samples.

Viewpoint Pairs	approach	Rank							
		5	10	15	20	25	30	35	40
$0°$ to $90°$	SDL-GS	**0.267**	**0.400**	**0.500**	**0.608**	**0.692**	**0.742**	**0.775**	**0.800**
	SDL-SC	0.242	0.367	0.450	0.525	0.608	0.692	0.750	0.783
	SVD	0.242	0.367	0.450	0.525	0.633	0.725	0.7667	0.7917
$0°$ to $180°$	SDL-GS	0.098	0.431	**0.627**	**0.706**	0.745	0.824	**0.863**	0.922
	SDL-SC	0.098	0.431	0.549	0.667	0.745	0.824	0.843	0.922
	SVD	0.098	0.412	0.588	0.667	**0.765**	0.804	0.843	**0.941**
$45°$ to $135°$	SDL-GS	0.056	**0.417**	**0.556**	0.611	0.694	**0.778**	**0.917**	–
	SDL-SC	0.028	0.361	0.500	0.583	0.694	0.722	0.889	–
	SVD	0.056	0.361	0.528	0.611	0.694	0.750	0.889	–

Comparison with Other Paradigms To validate the effectiveness of proposed approach on overall dataset, we compare with the LMNN [7] and ELF200 [6], which are the state-of-the-art approaches in VIPeR dataset. We use the implementation of LMNN in [23]. We also implement the ELF200 reported in [6]. We conduct the proposed approach on each subset of VIPeR dataset according to the viewpoints. The λ is set as 0.02 while the dictionary size is set as 250.

Fig. 2 shows the CMC curves of competing approaches. SDL-GS is observed to be superior or comparable than the supervised approach of metric learning. We emphasize that our approach is unsupervised and can be easily combined with supervised approaches to further achieve performance.
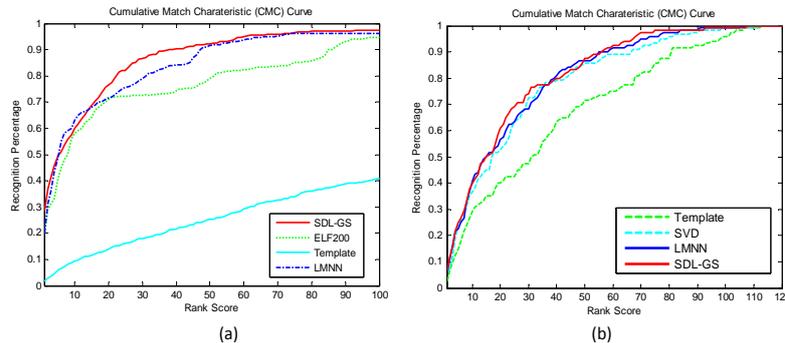
Fig. 2. Performance comparison among SDL-GS and other representative approaches. LMNN refers to the state-of-the-art approach reported in [7]. And ELF200 refers to [6]. Template is a baseline approach that directly compares distances between features of probe sample and gallery samples. (a) CMC curve comparison on all data; (b) CMC curve comparison on view pairs $0°$ to $90°$. Please refer to text (4) for detailed experimental setting.

5   Conclusion

In this work, we have presented a multi-view dictionary learning approach with shared latent embedding. Group sparsity constraints by $l_{1,\infty}$ regularization were adopted for more stable representation and superior performance. Our generative model provids a natural way to reconstruct data in unseen views. Such reconstruction ability is utilized in multi-view pedestrian and gait recognition. And various evaluation measures have demonstrated that SDL-GS outperforms other approaches in two mainstream datasets.

Currently, our model requires the knowledge of the viewpoint and that the training set should cover all viewpoints. In future work, we will investigate predicting viewpoint of the probe sample and training with incomplete views. We will also explore other extensions, such as by incorporating supervision information and learning Region of Interests (ROI).

References

1. Su, H., Sun, M., Fei-Fei, L., Savarese, S.: Learning a dense multi-view representation for detection, viewpoint classification and synthesis of object categories. In: ICCV. (2009) 213–220
2. Wu, B., Ram, N.: Cluster boosted tree classifier for multi-view, multi-pose object detection. In: ICCV. (2007) 1–8
3. Weinland, D., Boyer, E., Ronfard, R.: Action recognition from arbitrary views using 3d exemplars. In: ICCV. (2007) 1–7
4. Junejo, I., Dexter, E., Laptev, I.: Cross-view action recognition from temporal self-similarities. In: ECCV. (2008) 1–8
5. Prosser, B., Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: BMVC. (2010) 21.1–21.11
6. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: ECCV. (2008) 262–275
7. Dikmen, M., Akbas, E., Huang, T., Ahuja, N.: Pedestrian recognition with a learned metric. In: ACCV. (2010) 501–512

8. Bashir, K., Xiang, T., Gong, S.: Cross-view gait recognition using correlation strength. In: BMVC. (2010) 109.1–109.11
9. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: ECCV. (2006) 151–163
10. Kusakunniran, W., Q, W., Li, H., Zhang, J.: Multiple views gait recognition using view transformation model based on optimized gait energy image. In: ICCV workshop. (2009) 1058–1064
11. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. Trans. PAMI **31** (2009) 210–227
12. Yang, M., Zhang, L.: Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In: ECCV. (2010) 448–461
13. Jenatton, R., Obozinski, G., Bach, F.: Structured sparse principal component analysis. In: AISTATS. (2010) 366–373
14. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Ann. Stat. **32** (2004)
15. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: NIPS. (2007) 801–808
16. Huang, J., Huang, X., Metaxas, D.: Learning with dynamic group sparsity. In: ICCV. (2009) 64–71
17. Bengio, S., Pereira, F., Singer, Y., Strelow, D.: Group sparse coding. In: NIPS. (2009) 82–89
18. Chen, X., Pan, W., Kwok, J.T., Carbonell, J.G.: Accelerated gradient method for multi-task sparse learning problem. ICDM **0** (2009) 746–751
19. Jia, Y., Salzmann, M., Darrell, T.: Factorized latent spaces with structured sparsity. In: NIPS. (2010) 982–990
20. Nesterov, Y.: Gradient methods for minimizing composite objective function. Technical report, Euro (2007)
21. Vedaldi, A., Fulkerson, B.: Vlfeat – an open and portable library of computer vision algorithms. In: ACM Multimedia. (2010) 1469–1472
22. Ojala, T., Pietik1in, M., Harwood, D.: Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: ICPR. (2004) 582–585
23. Weinberger, K., Blitzer, J., Saul, L.: Distance metric learning for large margin nearest neighbor classification. In: NIPS. (2006) 1473–1480