Side-Information based Linear Discriminant Analysis for Face Recognition

Meina Kan[1,2,3]
mnkan@jdl.ac.cn
Shiguang Shan[1,2]
sgshan@jdl.ac.cn
Dong Xu[3]
dongxu@ntu.edu.sg
Xilin Chen[1,2]
xlchen@jdl.ac.cn

[1] Digital Media Research Center, Institute of Computing Technology, CAS, Beijing, China
[2] Key Laboratory of Intelligent Information Processing, Chinese Academy of Sciences, Beijing, China
[3] School of Computer Engineering, Nanyang Technological University, Singapore

Recently face recognition in unconstrained environment has attracted increasing attentions especially since the large scale unconstrained database Labeled Face in the Wild (LFW) [1] is released.

LFW has two different training modes: image-restricted mode and image-unrestricted mode. In the former mode, only *side-information*, *i.e*., whether a pair of images belongs to the same class (also referred as image pairs hereafter), is available, while in the latter mode, the full class label information is provided.

Most methods are evaluated under the image-restricted mode which only provides weakly-supervised side-information. Many of them deal with the side-information scenario by employing the typical two-class SVM classifier. However the multi-class methods including Fisherface [2] and its numerous extensions cannot be used in this scenario because the crucial class label information are not provided in the image-restricted evaluation mode.

In this work we propose a Side-Information based Linear Discriminant Analysis (SILD) method that can work well only with side-information, in which the within-class and between-class scatter matrices are computed by directly using the side-information. Specifically, the same-class image pairs are directly used to calculate the within-class scatter matrix and the different-class image pairs are employed to calculate the between-class scatter matrix.

It is worth mentioning that our method is different from the two-class FLDA, specifically only one projection direction can be obtained by using two-class Fisher Linear Discriminant Analysis (FLDA) while much more projection directions can be obtained by using our method. Moreover, we theoretically prove that, our SILD method is equivalent to multi-class FLDA when class labels are provided.

Let us denote $S=\{(x_i,x_j): l(x_i)=l(x_j)\}$ as the set of same-class image pairs and $D=\{(x_m,x_n): l(x_m)\neq l(x_n)\}$ as the set of different-class image pairs, with $l(x)$ representing the class label of image x. Then, the within-class and between-class scatter matrices can be respectively defined as follows:

$$S_W^{sild} = \sum_{(x_i,x_j)\in S} (x_i-x_j)(x_i-x_j)^T \quad (1)$$

$$S_B^{sild} = \sum_{(x_m,x_n)\in D} (x_m-x_n)(x_m-x_n)^T \quad (2)$$

Compared with FLDA, the new definition do not need know the identity of each sample and only use the weakly-supervised side-information to directly calculate the total within-class and between-class scatter matrices.

Similarly to FLDA, the projection matrix in SILD can be obtained by solving the following optimization problem:

$$W_{opt}^{sild} = \arg\max_W \frac{\left|W^T S_B^{sild} W\right|}{\left|W^T S_W^{sild} W\right|} \quad (3)$$

If the full label information is available, the equation (1) and (2) can be reformulated as:

$$S_W^{sild} = \sum_{i=1}^{c}\sum_{k=1}^{n_i}\sum_{l=1}^{n_i}(x_{ik}-x_{il})(x_{ik}-x_{il})^T = 2\sum_{i=1}^{c} n_i \sum_{k=1}^{n_i}(x_{ik}-m_i)(x_{ik}-m_i)^T \quad (4)$$

$$S_B^{sild} = 2rS_B - S_W^{sild} + 2rS_W \quad (5)$$

Given the new definition of within-class and between-class scatter matrices, the projection matrix of SILD can be solved as follows:

$$W_{opt}^{sild} = \arg\max_W \frac{\left|W^T S_B^{sild} W\right|}{\left|W^T S_W^{sild} W\right|} = \arg\max_W trace\left(\frac{W^T S_T W}{W^T S_W^{sild} W}\right) \quad (6)$$

If all classes have the same number of samples, we further have:

$$W_{opt}^{sild} = \arg\max_W trace\left(\frac{W^T S_T W}{2nW^T S_W W}\right) = \arg\max_W \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|} = W_{opt} \quad (7)$$

From the above equations, it is obvious that the projection matrix of SILD is identical to that of FLDA if the class label information is provided and all classes have the same number of samples. If each class has different number of samples, SILD is a variant of FLDA by focusing more on the classes with more samples (see the weight n_i in (4)). When the class label information is unavailable, SILD can be considered as an approximation of FLDA by exploiting a small fraction of full class label information only.

Inspired by [3], a more discriminative model can be learnt if the samples near the boundary are emphasized. However, the method in [3] cannot be directly used without the class label information. As in (1), the samples that are hard to be classified are already emphasized in the definition of S_W^{sild}. On the other hand, for S_B^{sild}, less attention is paid on the pair of samples that are more difficult. So, we reweight the pairs in S_B^{sild} to emphasize the samples that are hard for classification as follows:

$$\begin{aligned} S_B^{sild} &= \sum_{(x_m,x_n)\in D} w(x_m,x_n)(x_m-x_n)(x_m-x_n)^T \\ w(x_m,x_n) &= cosine(x_m,x_n) \end{aligned} \quad (8)$$

When the S_B^{sild} is calculated with (6), we refer to our method as 'Weighted SILD'.

The proposed method is evaluated on the large-scale unconstrained LFW database combined with Intensity, LBP, Gabor and Block-Gabor features respectively. After fusion of these features as shown in Fig.1, the proposed method can achieve comparable performance using less feature types when compared with the state-of-the-art result.

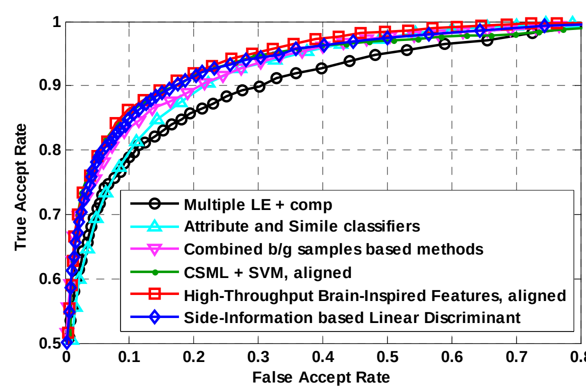

Figure 1: Performance of combined SILD and other state-of-the-art methods on the LFW database under image-restricted protocol.

[1]. G.B. Huang, Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. in Technical Report, 2007.

[2]. P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 19(7):711-720, 1997.

[3]. Z. Li, D. Lin, and X. Tang. Nonparametric Discriminant Analysis for Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 31(4):755-761, 2009.