Discriminative Transfer Learning for Single-Sample Face Recognition

Junlin Hu[1], Jiwen Lu[2*], Xiuzhuang Zhou[3], Yap-Peng Tan[1]
[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Advanced Digital Sciences Center, Singapore
[3]College of Information Engineering, Capital Normal University, Beijing, China
jhu007@e.ntu.edu.sg, jiwen.lu@adsc.com.sg, xiuzhuang.zhou@cnu.edu.cn, eyptan@ntu.edu.sg

Abstract

Discriminant analysis is an important technique for face recognition because it can extract discriminative features to classify different persons. However, most existing discriminant analysis methods fail to work for single-sample face recognition (SSFR) because there is only a single training sample per person such that the within-class variation of this person cannot be estimated in such scenario. In this paper, we present a new discriminative transfer learning (DTL) approach for SSFR, where discriminant analysis is performed on a multiple-sample generic training set and then transferred into the single-sample gallery set. Specifically, our DTL learns a feature projection to minimize the intra-class variation and maximize the inter-class variation of samples in the training set, and minimize the difference between the generic training set and the gallery set, simultaneously. Experimental results on three face datasets including the FERET, CAS-PEAL-R1, and LFW datasets are presented to show the efficacy of our method.

1. Introduction

Face recognition has been widely investigated in biometrics due to its promising applications such as human computer interaction and surveillance. While a number of face recognition algorithms have been proposed [16, 10, 11, 12], most of them fail to work for single-sample face recognition (SSFR) because there is only a single training sample available for each person and the within-class variation cannot be estimated in such scenario. In many practical applications such as e-passport identification and law enhancement, there is usually a single training sample per individual in the system and how to extract discriminative information from such single training sample set remains a challenging problem in SSFR.

In recent years, many approaches have been proposed to figure out the SSFR problem [16, 21, 23, 20], and they can be roughly classified into three categories: feature learning only using the single gallery sample set, virtual sample generation, and feature learning with an auxiliary dataset. For the first category, the single sample gallery set is used to extract features. Principal component analysis (PCA) [17] is the most representative method, which seeks a low-dimensional subspace by maximizing the total scatter matrix of training samples. Many extensions of PCA have also been proposed, such as projection-combined PCA ($(\mathrm{PC})^2\mathrm{A}$) [19] and singular value decomposition (SVD) perturbation based PCA (SVD-PCA) [22]. $(\mathrm{PC})^2\mathrm{A}$ combines the original image and a new synthesized image by projecting the original one onto the horizontal and vertical directions, and performs PCA on the projection-combined image. Compared with PCA, $(\mathrm{PC})^2\mathrm{A}$ is robust to variations of illumination and expression. For SVD-PCA, it adapts the similar idea of $(\mathrm{PC})^2\mathrm{A}$, and their difference is that SVD-PCA obtains the enriched image by employing singular value perturbing technique on the original image. Recently, sparse representation based classification (SRC) [18] has been employed to address SSFR problem and has achieved competitive recognition performance with previous methods. However, these methods are unsupervised and discriminative information has not been exploited.

For the second class, multiple virtual samples are produced for each person so that the conventional linear discriminant analysis (LDA) [2] method is applicable to the problem of SSFR. The Block-LDA [3] is a simple and direct work, which firstly divided one whole image into some blocks with same size, and then treated these blocks as multiple samples per person to train LDA model. Lu *et al*. [11] also partitioned each image into several sub-images to generate multiple samples for each person, and employed a discriminative multi-manifold analysis (DMMA) method on these sub-images for learning discriminative features by maximizing the manifold margins of different persons as well as considering the geometrical information of the local blocks. Li *et al*. [9] introduced a variant of LDA

*Corresponding author.

named ensemble of randomized linear discriminant analysis (ERLDA) to generate extra training samples on an low-dimensional subspace by random projection and ensemble learning. Different from these methods by dividing an image into multiple blocks, SVD-LDA [4] approach decomposes the whole image into two components by SVD technique: a general appearance image and a difference image, where this difference image is utilized to calculate the within-class scatter matrix of this person and further extract discriminant features under the LDA framework. However, these methods usually need some prior knowledge to guide the generation of virtual samples, and it is not easy to guarantee the quality of virtual samples because there are usually high correlation among the generated virtual samples.

Regarding the third class, an additional generic set with multiple samples per subject is adopted to learn a discriminative model. The main assumption of such learning methods using generic training set is that faces of various subjects look alike such that they should have similar intra-class variation. Based on this assumption, the within-class scatter matrix of each subject in the gallery set can be estimated using a generic set, then LDA and its extensions can be utilized to address SSFR. In [8], Kim and Kitter proposed to learn a linear discriminant model from the generic training set to solve the single-sample face recognition. Recently, adaptive generic learning (AGL) [15] method was introduced to learn a generic discriminant model to address the SSFR problem. The AGL method can infer the within-class scatter matrix of single simple by linearly combining all the within-class scatter matrixes of subjects with multiple samples in the generic training set. Furthermore, adaptive discriminant analysis (ADA) [7] as an extension of AGL was proposed to calculate the within-class scatter of each subject in the gallery set by adopting the within-class scatters of several nearest neighbors persons in the generic set, then LDA is used and it achieves better performance than others. However, most methods in this category ignore the distribution difference between generic set and gallery set.

In this paper, we propose a discriminative transfer learning (DTL) method for SSFR, where discriminant analysis is performed on a generic training set with multiple samples per subject and transferred into a gallery set where each person has a single gallery image. Specifically, our DTL learns a projection to minimize the intra-class variation and maximize the inter-class variation of samples in the training set, as well as minimize the difference between the generic set and the gallery set, simultaneously. Experimental results on three face datasets show the efficacy of our proposed method.

2. Proposed Approach

This section firstly gives a brief introduction to our basic idea, then formulates the proposed DTL method.

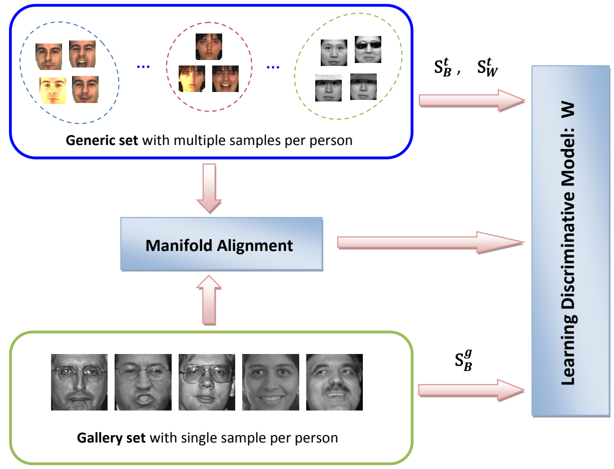

Figure 1. Intuitive illustration of the proposed DTL method. The DTL learns a discriminative model $\mathbf{W}$ by transferring the intra-class variation of the generic set to that of the gallery set, and minimizing the difference between the generic and gallery sets in a joint latent space via manifold alignment, simultaneously.

2.1. Basic Idea

Figure 1 shows the basic idea of the proposed DTL method on addressing the SSFR problem. Given the generic set ($\mathbf{X}^t$) including multiple samples for each person and the gallery set ($\mathbf{X}^g$) containing a single sample per person, respectively. The proposed DTL approach can learn a discriminative transfer model by maximizing the inter-class variation and minimizing the intra-class variation of samples in the training set, and simultaneously minimizing the distribution difference between these two datasets in the joint latent subspace via manifold alignment strategy. Finally, the learned discriminative model $\mathbf{W}$ is used to recognize the face from the single gallery set.

2.2. Notation

Let $\mathbf{X} = \{\mathbf{X}_c \in \mathbb{R}^{d \times N_c}\}_{c=1}^{C}$ be a set of samples, where $\mathbf{X}_c$ means samples from the class c; d is size of each sample; N_c is the number of samples belonging to class c; and C is the number of classes. In the linear discriminant analysis, the within-class scatter matrix $\mathbf{S}_W$ measures intra-class compactness, and the between-class scatter $\mathbf{S}_B$ measures the inter-class separability, and they can be computed as:

$$\mathbf{S}_W = \frac{1}{C}\sum_{c=1}^{C}\mathbf{S}_c, \tag{1}$$

$$\mathbf{S}_c = \frac{1}{N_c}\sum_{\mathbf{x}\in\mathbf{X}_c}(\mathbf{x}-\mathbf{m}_c)(\mathbf{x}-\mathbf{m}_c)^T, \tag{2}$$

$$\mathbf{S}_B = \frac{1}{C}\sum_{c=1}^{C}(\mathbf{m}_c-\mathbf{m})(\mathbf{m}_c-\mathbf{m})^T, \tag{3}$$

where $\mathbf{m}_c$ is the mean of the c-th class; matrix $\mathbf{S}_c$ denotes the within-class scatter of the class c; and vector $\mathbf{m}$ denotes the mean of all the samples.

2.3. Discriminative Transfer Learning

Denote by $\mathbf{X}^t = \{\mathbf{X}_c^t \in \mathbb{R}^{d\times N_c^t}\}_{c=1}^{C^t}$ the dataset of the generic set, and $\mathbf{X}^g = \{\mathbf{X}_c^g \in \mathbb{R}^{d\times N_c^g}\}_{c=1}^{C^g}$ the dataset of the gallery set, where superscript t and g mean samples from generic set (or training set) and gallery set, respectively; $N^g = \sum_{c=1}^{C^g} N_c^g$, $N^t = \sum_{c=1}^{C^t} N_c^t$. To transfer more information from generic set to gallery set, we seek a joint projection matrix $\mathbf{W} \in \mathbb{R}^{d\times r}$, $r < d$ by maximizing the following objective function:

$$\begin{aligned}\max_{\mathbf{W}}\ \mathcal{F}(\mathbf{W}) &= \mathrm{tr}\Big(\mathbf{W}^T(\mathbf{S}_B^t - \mu\,\mathbf{S}_W^t)\mathbf{W}\Big)\\ &\quad + \alpha\,\mathrm{tr}(\mathbf{W}^T\mathbf{S}_B^g\mathbf{W}) - \beta\,dist(\mathbf{X}^g,\mathbf{X}^t), \quad (4)\end{aligned}$$

where μ, α, and β are three positive parameters; and the term $dist(\mathbf{X}^g,\mathbf{X}^t)$ is the distance between manifolds $\mathbf{X}^g$ and $\mathbf{X}^t$, which can be computed in the transformed subspace:

$$\begin{aligned}dist(\mathbf{X}^g,\mathbf{X}^t) &= \frac{1}{N^gN^t}\sum_{i=1}^{N^g}\sum_{j=1}^{N^t}\|\mathbf{W}^T\mathbf{x}_i^g - \mathbf{W}^T\mathbf{x}_j^t\|^2\mathbf{A}_{ij}\\ &= \mathrm{tr}(\mathbf{W}^T\mathbf{H}\mathbf{W}), \quad (5)\end{aligned}$$

where matrix $\mathbf{H}$ is defined as:

$$\mathbf{H} = \frac{1}{N^gN^t}\sum_{i=1}^{N^g}\sum_{j=1}^{N^t}(\mathbf{x}_i^g - \mathbf{x}_j^t)(\mathbf{x}_i^g - \mathbf{x}_j^t)^T\mathbf{A}_{ij}, \quad (6)$$

and $\mathbf{A}_{ij}$ is an affinity matrix used to measure the similarity between $\mathbf{x}_i^g$ and $\mathbf{x}_j^t$, it can be calculated by heat kernel:

$$\mathbf{A}_{ij} = \exp(-\|\mathbf{x}_i^g - \mathbf{x}_j^t\|^2/\sigma^2). \quad (7)$$

By substituting Equation (5) into Equation (4), we have the following formulation:

$$\begin{aligned}&\max_{\mathbf{W}}\ \mathcal{F}(\mathbf{W}) =\\ &\quad\mathrm{tr}\Big(\mathbf{W}^T(\mathbf{S}_B^t - \mu\,\mathbf{S}_W^t + \alpha\,\mathbf{S}_B^g - \beta\,\mathbf{H})\mathbf{W}\Big). \quad (8)\end{aligned}$$

Then, $\mathbf{W}$ can be obtained by gradient descent method:

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \eta\,\nabla\mathcal{F}(\mathbf{W}_k), \quad (9)$$

where η is the learning rate, $\mathbf{W}_k$ is the k-th iteration of $\mathbf{W}$, and $\nabla\mathcal{F}(\mathbf{W})$ is the gradient with respect to $\mathbf{W}$, which can be expressed by:

$$\nabla\mathcal{F}(\mathbf{W}) = 2(\mathbf{S}_B^t - \mu\,\mathbf{S}_W^t + \alpha\,\mathbf{S}_B^g - \beta\,\mathbf{H})\mathbf{W}. \quad (10)$$

Lastly, nearest neighbor (NN) classifier is employed to recognize single face. The proposed DTL method is summarized in Algorithm 1, where $\mathbf{I}_{d\times r}$ is a matrix with 1s on the diagonal and zeros elsewhere.

Algorithm 1: DTL

Input: The generic training set $\mathbf{X}^t$; The single gallery set $\mathbf{X}^g$; Positive constants: μ, α, β; The learning rate η; Total iterative number T; Convergence error ε.

Output: Projection matrix $\mathbf{W}$.

```
Calculating S^t_W, S^t_B, S^g_B, and H.
Initialing k = 0, W_0 = I_{d×r}.
Calculating F(W_0).
while k < T do
    k ← k + 1.
    Computing W_k by using Eqs. (9) and (10).
    Calculating F(W_k).
    If |W_k − W_{k−1}| < ε, then go to Return.
end while
Return: W ← W_k.
```

3. Experiments

To evaluate the effectiveness of our DTL approach, we perform experiments on three widely used datasets: FERET [14], CAS-PEAL-R1 [5], and LFW databases [6]. We also compare our DTL method with several popular approaches on addressing the challenge of SSFR.

3.1. Databases

FERET database. The Face Recognition Technology (FERET) [14] database totally covers face images of 1196 different persons with the variations such as lighting, expression, and aging. Following the standard partitions of this dataset, there are 6 subsets for different tasks: *training set*, *gallery set*, *fb*, *fc*, *dup1* and *dup2*, respectively, and *fb*, *fc*, *dup1* and *dup2* are selected as four probe sets. The *training set* contains 1002 images from 429 individuals; the *gallery set* includes 1196 images of 1196 persons, one image per person; there are 1195 images in the probe set *fb* captured under another facial expression; the *fc* consists of 194 images with different lighting; the *dup1* has 722 face images captured from different period; and there are 234 images in the *dup2*, which is a subset of the *dup1* and over the period of one year.

CAS-PEAL-R1 database. CAS-PEAL-R1 [5] is the largest Chinese face database for training and evaluating face recognition methods, where the face images were taken with various variations such as expression, accessory, lighting and pose. Following the standard settings, there are three kinds of datasets in this database: database: training set, gallery set, and probe set. The training set consists of 1200 images of 300 subjects, 4 images for each individual. The gallery set contains 1040 images from 1040 subjects with a normal condition. For the probe set, three subsets are selected in our experiments: accessory, expression, lighting

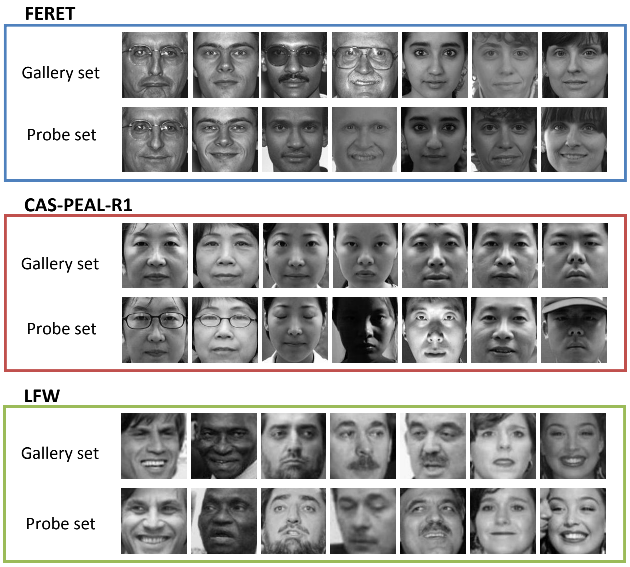

Figure 2. Some cropped face examples from FERET, CAS-PEAL-R1, and LFW databases, respectively.

subsets, and there are 2285, 1570 and 2243 images corresponding to 438, 377 and 233 subjects, respectively.

LFW database. The Labeled Face in the Wild (LFW) dataset [6] contains 13233 labeled face images of 5749 people, where 1680 individuals of them have two or more face images. Images in this dataset are taken in uncontrolled settings, including variations in pose, scale, lighting, background, hairstyle, expression, partial occlusion, resolution, and so on, which poses a great challenge to our recognition task. In our experiments, we take 3360 images from 1680 persons with 2 images per person for conducting experiments, for those subjects whose images are more than two, we just select the first two images, and one falls into gallery set, another falls into probe set. Figure 2 shows some cropped face examples from these three databases.

3.2. Experimental Settings

To validate our method, the generic training set with multiple samples per person was collected from two public face databases: AR database [13] and CAS-PEAL-R1 database. The AR database contains 135 subjects, where each subject has several images up to 26, with different illumination conditions, expressions, and facial disguises (wearing sunglasses and wearing scarf). We discard the images with facial disguises, finally we totally gather 1645 images from these 135 people where each person has multiple samples. For the CAS-PEAL-R1 database, all the images in the training set, namely 1200 frontal facial images from 300 subjects, are used to build up the generic training set. Therefore, there are 2845 images from 435 persons in the generic training set. Figure 3 lists some examples from the generic set used in our experiments.

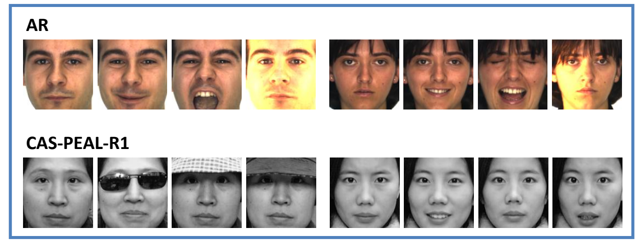

Figure 3. Some cropped face examples from the generic set.

In addition, all the face images are aligned and normalized to the size of 64×64 in line with the eye coordinates, furthermore, each image is reduced to 400 dimension by P-CA where the projection is trained on the generic training set. In addition, the free parameters μ, α, β and the learning rate η in our method were empirically set as 1, 1, 0.01 and 10^{-4}, respectively. Finally, the nearest neighbor classifier with Euclidean distance metric is used to recognize face from the gallery set with single simple per person.

3.3. Results and Analysis

Comparison with Existing SSFR Methods: We compare our DTL approach with some typical methods that are often utilized to address the SSFR problem. These methods include PCA [17], $(\mathrm{PC})^2$A [19], SVD-PCA [22], SVD-LDA [4], Block-LDA [3], Generic-LDA, adaptive generic learning (AGL) [15], adaptive discriminant analysis (ADA) [7], and ensemble of randomized linear discriminant analysis (ERLDA) [9]. For fair comparison, we also implemented these methods except AGL and ERLDA by ourselves. For PCA, it is trained on the single gallery set to generate projection matrix. For $(\mathrm{PC})^2$A and SVD-PCA, the weighting parameter α was set as 0.25 according to suggestion in [19, 22], respectively. As for SVD-LDA, the first three largest singular values were used to generate a virtual sample to estimate the within-class scatter matrix. For Block-LDA, the whole image is evenly divided into several blocks with size of 10×10 to construct multiple virtual samples. For Generic-LDA, LDA was performed on the generic training set. As for ADA, the number of K-nearest neighbor was set to 10. For these methods, the nearest neighbor classifier with Euclidean distance metric is used for recognition, and the best result is recorded by testing all dimensions.

Tables 1-3 tabulate some comparison results of our method versus other methods on the FERET, CAS-PEAL-R1, and LFW databases, respectively. From these tables, we can observe that DTL consistently outperforms other compared method in recognition accuracy, this observation shows that our proposed is effective to solve single-sample face recognition problem. Furthermore, we also have some following observations: 1) Recognition methods only using the single gallery set, such as PCA, $(\mathrm{PC})^2$A, SVD-PCA and SVD-LDA, show poor performance on the most tests;

Table 1. The comparison of rank-1 recognition accuracy (%) on the FERET database. Here, notation * means that the results are taken from the original paper.

Method	*fb*	*fc*	*dup1*	*dup2*
PCA	84.52	22.16	39.47	13.25
$(\mathrm{PC})^2\mathrm{A}$	83.93	15.46	38.64	11.54
SVD-PCA	83.10	10.82	37.81	10.68
SVD-LDA	85.02	26.80	37.40	14.53
Block-LDA	83.35	65.46	42.80	41.03
Generic-LDA	86.95	71.65	50.28	40.60
ADA	87.62	73.71	53.88	41.03
DTL	**89.79**	**76.80**	**55.96**	**44.44**
AGL* [15]	88.50	71.60	53.30	35.00
ADA* [7]	**90.10**	74.80	52.50	36.80
ERLDA* [9]	**92.40**	70.80	52.40	38.00

Table 2. The comparison of rank-1 recognition accuracy (%) on the CAS-PEAL-R1 database.

Method	Accessory	Expression	Lighting
PCA	40.26	78.54	6.06
$(\mathrm{PC})^2\mathrm{A}$	38.60	78.09	4.86
SVD-PCA	36.50	76.05	4.01
SVD-LDA	41.27	77.83	6.29
Block-LDA	38.73	77.64	12.88
Generic-LDA	59.56	82.48	19.08
ADA	60.96	82.87	22.65
DTL	**62.45**	**85.60**	**25.46**

Table 3. The comparison of rank-1 recognition accuracy (%) on the LFW database.

Method	Accuracy (%)
PCA	7.44
$(\mathrm{PC})^2\mathrm{A}$	6.90
SVD-PCA	6.49
SVD-LDA	6.73
Block-LDA	7.44
Generic-LDA	9.05
ADA	10.36
DTL	**11.79**

2) The Generic-LDA gives better performance on the three databases, which demonstrates that the generic set can approximate the within-class variation of the single sample to some extent; 3) The ADA performs as well as the Generic-LDA approach; and 4) The proposed DTL method achieves significant improvement than methods based on the single gallery set with large margin, and consistently performs better than ADA through all the experiments. This illustrates that our method can transfer more discriminative information from the generic training set to the single gallery set by manifold alignment strategy.

Another important observation is that the recognition

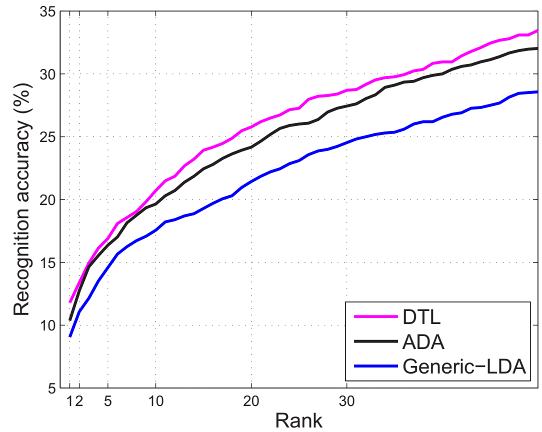

Figure 4. The recognition accuracy versus various ranks on the LFW database.

Table 4. The rank-1 recognition accuracy (%) of two different features on the FERET database.

Feature	*fb*	*fc*	*dup1*	*dup2*
Intensity	89.79	76.80	55.96	44.44
LBP	**95.82**	**96.39**	**70.91**	**70.09**

Table 5. The rank-1 recognition accuracy (%) of two different features on the CAS-PEAL-R1 database.

Feature	Accessory	Expression	Lighting
Intensity	62.45	85.60	25.46
LBP	**82.58**	**92.93**	**32.46**

performance of all the methods on the real-world LFW database are much lower than those on the FERET and CAS-PEAL-R1 databases. Figure 4 also plots the recognition accuracy versus different ranks of the three methods on the LFW database. We can find the recognition rates according to rank-30 are still below 30%, which implies that face recognition are very challenging on the real-world images due to large variation on scale, pose, lighting, background, hairstyle, expression, and so on.

Performance with / without Local Features: We also tested our method using local binary patterns (LBP) [1] feature. Each image is first partitioned into 8×8 non-overlapping blocks, then a 59-bin uniform pattern LBP is computed for each block, lastly these LBPs are concatenated to form a long vector. Tables 4 and 5 show the recognition accuracy of the proposed method using intensity and LBP features on two databases, respectively. We can see our method with local LBP feature consistently outperforms intensity in a large gain, and these results show that local LBP feature is more robust for single sample face recognition.

Parameter Analysis: Figure 5 shows the rank-1 recognition accuracy of the proposed DTL method with various value of parameters α and β on the the LFW database. We

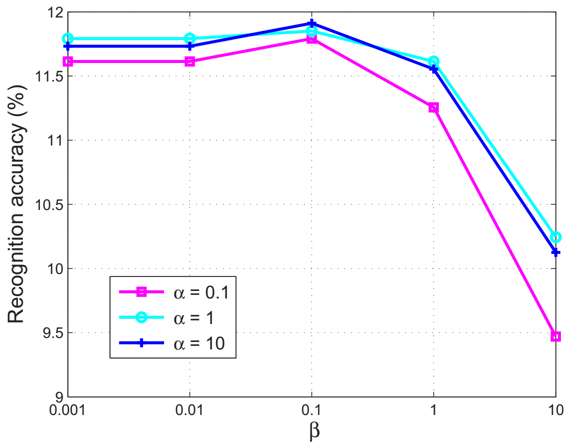

Figure 5. The recognition accuracy of the DTL method with various parameters (α and β) on the LFW database.

can observe that our method is robust to changes of parameters α and β at a large range.

4. Conclusion

In this paper, we have proposed a discriminative transfer learning (DTL) method to address the singe-sample face recognition problem. Our DTL approach transfers the intra-personal variation of the generic training set containing multiple samples per subject to that of the gallery set with single sample per person, and simultaneously aligns these two datasets in a joint latent space via manifold alignment. Therefore, it can learn more discriminative information for identification. Experimental results on several face datasets have shown the efficacy of the proposed method.

Acknowledgement

This work is partly supported by a research grant from the Agency for Science, Technology and Research (A*STAR) of Singapore for the Human Centric Cyber Systems (HCCS) Program at the Advanced Digital Sciences Center (ADSC), and the National Natural Science Foundation of China under Grant 61373090.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[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, July 1997.

[3] S. Chen, J. Liu, and Z.-H. Zhou. Making flda applicable to face recognition with one sample per person. *Pattern Recognition*, 37(7):1553–1555, 2004.

[4] Q. Gao, L. Zhang, and D. Zhang. Face recognition using flda with single training image per person. *Applied Mathematics and Computation*, 205(2):726–734, 2008.

[5] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The cas-peal large-scale chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 38(1):149–161, 2008.

[6] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[7] M. Kan, S. Shan, Y. Su, D. Xu, and X. Chen. Adaptive discriminant learning for face recognition. *Pattern Recognition*, 46(9):2497–2509, 2013.

[8] T.-K. Kim and J. Kittler. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):318–327, 2005.

[9] Y. Li, W. Shen, X. Shi, and Z. Zhang. Ensemble of randomized linear discriminant analysis for face recognition with single sample per person. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–8, 2013.

[10] J. Lu and Y.-P. Tan. Regularized locality preserving projections and its extensions for face recognition. *IEEE Transactions on Systems, Mans, and Cybernetics, Part B: Cybernetics*, 40(3).

[11] J. Lu, Y.-P. Tan, and G. Wang. Discriminative multimanifold analysis for face recognition from a single training sample per person. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):39–51, 2013.

[12] J. Lu, G. Wang, W. Deng, and P. Moulin. Simultaneous feature and dictionary learning for image set based face recognition. In *European Conference on Computer Vision*, pages 265–280, 2014.

[13] A. M. Martinez and R. Benavente. The ar face database. Technical Report 24, CVC Technical Report, June 1998.

[14] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image Vision Computing*, 16(5):295–306, 1998.

[15] Y. Su, S. Shan, X. Chen, and W. Gao. Adaptive generic learning for face recognition from a single sample per person. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2699–2706, 2010.

[16] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725–1745, 2006.

[17] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[18] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[19] J. Wu and Z.-H. Zhou. Face recognition with one training image per person. *Pattern Recognition Letters*, 23(14):1711–1719, 2002.

[20] H. Yan, J. Lu, X. Zhou, and Y. Shang. Multi-feature multi-manifold learning for single-sample face recognition. *Neurocomputing*, 143:134–143, 2014.

[21] M. Yang, L. J. V. Gool, and L. Zhang. Sparse variation dictionary learning for face recognition with a single training sample per person. In *IEEE International Conference on Computer Vision*, pages 689–696, 2013.

[22] D. Zhang, S. Chen, and Z.-H. Zhou. A new face recognition method based on svd perturbation for single example image per person. *Applied Mathematics and Computation*, 163(2):895–907, 2005.

[23] L. Zhuang, A. Y. Yang, Z. Zhou, S. S. Sastry, and Y. Ma. Single-sample face recognition with image corruption and misalignment via sparse illumination transfer. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3546–3553, 2013.