Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Transfer subspace learning for cross-dataset facial expression recognition

Haibin Yan

School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China

ARTICLE INFO

Article history:
Received 25 September 2015
Received in revised form
21 November 2015
Accepted 21 November 2015

Keywords:
Facial expression recognition
Transfer learning
Subspace learning
Multi-view learning
Biometrics

ABSTRACT

In this paper, we propose a transfer subspace learning approach cross-dataset facial expression recognition. To our best knowledge, this problem has been seldom addressed in the literature. While many facial expression recognition methods have been proposed in recent years, most of them assume that face images in the training and testing sets are collected under the same conditions so that they are independently and identically distributed. In many real applications, this assumption does not hold as the testing data are usually collected online and are generally more uncontrollable than the training data. Hence, the testing samples are likely different from the training samples. In this paper, we define this problem as cross-dataset facial expression recognition as the training and testing data are considered to be collected from different datasets due to different acquisition conditions. To address this, we propose a transfer subspace learning approach to learn a feature subspace which transfers the knowledge gained from the source domain (training samples) to the target domain (testing samples) to improve the recognition performance. To better exploit more complementary information for multiple feature representations of face images, we develop a multi-view transfer subspace learning approach where multiple different yet related subspaces are learned to transfer information from the source domain to the target domain. Experimental results are presented to demonstrate the efficacy of these proposed methods for the cross-dataset facial expression recognition task.

1. Introduction

Over the past decades, facial expression analysis has been extensively studied in the area of pattern recognition and computer vision due to its huge potential applications such as intelligent affective computing and human robot interaction. Over the past two decades, a number of facial expression analysis methods [1–4] have been proposed in the literature. Generally, there are two key components in a facial expression recognition system: feature representation and expression classification. For feature representation, there are mainly two categories of methods: geometry-based and appearance-based. Geometry-based methods usually extract facial features such as the shape and locations of facial components (like the mouth, eyes, brows and nose) and represent them by a feature vector to characterize the facial geometry [5–8]. In general, different facial expressions have different feature representations. Appearance-based methods holistically convert each facial image into a feature vector and then apply subspace learning techniques to extract some statistical features for facial expression representation [9–15]. Since it is generally challenging to precisely localize and extract geometrical features for geometry-based methods in many practical applications, appearance-based methods are more popular for facial expression recognition and also demonstrate better performance than geometry-based ones in terms of the recognition accuracy. Having obtained feature representations of facial images, facial expression classification can be implemented by using a multi-class classification method, such as the nearest neighborhood classifier, support vector machine, neural networks and sparse representation. See [1–4] for more details.

Subspace learning techniques have been widely used to reveal the intrinsic structure of data and successfully applied to facial expression recognition. With subspace learning methods [39–58], facial expression images are projected into a low-dimensional feature space to reduce the feature dimensions. Representative methods in subspace learning include principal component analysis (PCA) [16,29], linear discriminant analysis (LDA) [17,29], locality preserving projections (LPPs) [18,30] and orthogonal neighborhood preserving projections (ONPPs) [19]. PCA [16,29] aims to learn a feature subspace by mapping the original high-dimensional face images to a low-dimensional linear subspace which is spanned by the top eigenvectors of a covariance matrix. LDA [17,29] seeks a discriminative subspace which consists of a set

E-mail address: eyanhaibin@bupt.edu.cn

http://dx.doi.org/10.1016/j.neucom.2015.11.113
0925-2312/

Please cite this article as: H. Yan, Transfer subspace learning for cross-dataset facial expression recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.11.113

of projections to maximize the ratio of the between-class variance to the within-class variance. Generally, LDA usually outperforms PCA for classification because it exploits the label information of samples. Unlike PCA and LDA which fail to discover the underlying nonlinear manifold structure of face samples, a number of manifold-based subspace learning algorithms [16,18,29,30,37,38] have been proposed in recent years and some of them have been successfully employed for facial expression recognition. The basic idea of these methods is high-dimensional face images can be modeled as a set of geometrically related points which lie on or nearby a smooth and low-dimension manifold. LPP is the most representative manifold-based subspace learning method, where it seeks a low-dimensional subspace to preserve the intrinsic geometric structure of the original samples. Since the bases of LPP are not orthogonal, ONPP is introduced to enforce an orthogonal constraint on the optimization objective of LPP.

Recently, Shan et al. [9] compared these methods for facial expression recognition and reported that supervised LPP was the best one in supervised methods and OPPP produced the best results in unsupervised methods. More recently, Xiao et al. [20] proposed a multi-manifold learning method for facial expression recognition, in which each expression data was modeled by a manifold and the recognition was performed by using a “data-to-manifold” distance strategy. Experimental results on two benchmark face datasets have shown the advantage of their proposed method. While different subspace learning methods are developed with different motivations, they can be well interpreted into a general graph embedding framework [30].

Most existing facial expression recognition methods assume facial images in the training and testing sets are collected under the same condition such that they are independent and identically distributed. In many real world applications, this assumption may not hold as the testing data are usually collected online and generally more uncontrollable than the training data, such as different races, illuminations and imaging conditions. Under this scenario, the performance of conventional subspace learning methods may be poor because the training and testing data are not independently and identically distributed. The generalization capability of these methods is limited on the cross-dataset facial expression recognition problem. To the best of our knowledge, this problem has not been formally addressed in the literature even if it is very important in many real applications.

To address this problem, we propose a new transfer subspace learning approach to learn a feature space which transfers the knowledge gained from the training set to the target (testing) data to improve the recognition performance under cross-dataset scenarios. We apply the proposed approach to four popular subspace learning methods including PCA, LDA, LPP and ONPP, and formulate the corresponding transfer PCA (TPCA), transfer LDA (TLDA), transfer LPP (TLPP) and transfer ONPP (TONPP) for cross-dataset facial expression recognition. To better exploit more complementary information for multiple feature representations of face images, we develop a multi-view transfer subspace learning approach where multiple different yet related subspaces are learned to transfer information from the source domain to the target domain, where four multi-view transfer subspace learning methods, namely multi-view transfer PCA (MTPCA), multi-view transfer LDA (MTLDA), multi-view transfer LPP (MTLPP), and multi-view transfer ONPP (MTONPP) are introduced. Experimental results are presented to demonstrate the efficacy of the proposed methods for cross-dataset facial expression recognition.

This paper is an extended version of work presented at IEEE ICRA 2011 [31]. New contributions include the newly proposed localized multi-view transfer subspace learning methods, analysis of the proposed approach, and extensive experimental results.

2. Related work

In this section, we briefly review three related topics: subspace learning, transfer learning, and multi-view learning.

2.1. Subspace learning

Let $X=[x_1,x_2,\dots,x_N]$, $x_i \in R^d$, $i=1,2,\dots,N$, be a training set of facial images, where N is the number of samples and d is the feature dimension of each sample. For supervised subspace learning algorithms, the class label of x_i is assumed to be $l_i \in \{1,2,\dots,c\}$, where c is the number of classes. For the jth class, n_j denotes the number of its samples, where $j=1,2,\dots,c$. Hence, $N=\sum_{j=1}^{c} n_j$. The objective of a subspace learning algorithm, such as PCA, LDA, LPP and ONPP, is to find a linear projection matrix $W=[w_1,w_2,\dots,w_k]$ to map x_i into a low dimensional representation y_i, where $y_i = W^T x_i \in R^m$, $m<d$ [9]. The essential differences of different subspace learning methods lie in their differences in defining and finding the projection matrix W by using different objective functions and constraints, such as

$$\min \quad F(W) \text{ subject to } \quad G(W)=0 \tag{1}$$

Table 1 shows the objective functions and constraints of PCA, LDA, LPP and ONPP, where $S_T=\frac{1}{N}\sum_{i=1}^{N}(x_i-m)(x_i-m)^T$, $m=\frac{1}{N}\sum_{i=1}^{N} x_i$, $S_B=\frac{1}{N}\sum_{i=1}^{c} n_i(m_i-m)(m_i-m)^T$, $S_W=\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_{ij}-m_i)(x_{ij}-m_i)^T$, x_{ij} denotes the jth training sample of the ith class, m_i is the mean of the training samples of the ith class, $L=D-S$, $D_{ii}=\sum_j S_{ji}$, S_{ij} is the locality similarity between x_i and x_j, $M=(I-V^T)(I-V)$, V can be obtained by solving the following optimization function:

$$\min \varepsilon(V)=\sum_i \| x_i - \sum_k V_{ik} x_{ik} \|^2 \tag{2}$$

where x_{ik} is the k-nearest neighbors of x_i.

2.2. Transfer learning

The past five years have witnessed the significance of transfer learning for practical applications such as cross-domain image and text classification, and domain-adaptation video analysis. It has also been identified to be an effective solution to address the cross-dataset recognition problem because it can transfer the knowledge gained from the training set to the testing set. Generally, there are three main issues in transfer learning: what to transfer, how to transfer and when to transfer. Compared with the conventional machine learning techniques, transfer learning can be generally classified into three categories: inductive transfer learning, transductive transfer learning and unsupervised transfer learning. Refer to [21] for more details.

While a number of transfer learning methods have been proposed recently, there is little effort of transfer learning made for subspace learning. To our knowledge, Si et al. [22] first applied transfer learning techniques to subspace learning by minimizing the distribution distance between the source and target domains in subspace learning algorithms. More recently, Su et al. [23] employed the mixture Gaussian model to model the distributions of the data

Table 1
Objective functions and constraints of four popular subspace learning methods.

Method	$F(W)$	$G(W)$
PCA	$-\mathrm{tr}(W^T S_T W)$	$W^T W - I$
LDA	$\frac{\mathrm{tr}(W^T S_W W)}{\mathrm{tr}(W^T S_B W)}$	-
LPP	$W^T X L X^T W$	$W^T X D X^T W - I$
ONPP	$\mathrm{tr}(W^T X M X^T W)$	$W^T W - I$

Please cite this article as: H. Yan, Transfer subspace learning for cross-dataset facial expression recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.11.113

in the source and target domains to make it more consistent with the original LDA method. However, these methods estimate the distribution based on the kernel density estimation method and Gaussian model, respectively, which may fail when there is a limited number of samples in the source and target domains. In this paper, we propose a new nonparametric transfer learning approach to learn a feature space which transfers the knowledge gained from the training set to the target (testing) data to improve the recognition performance under cross-dataset scenarios.

2.3. Multi-view learning

In many real-world applications, samples are usually characterized as a multi-view data representation. For a given face image, we extract many different descriptors such as intensity, local binary patterns (LBPs) and Gabor features. Generally, different features characterize samples from different perspectives, and the presence of multi-view data provide an opportunity to learn better representations for recognition tasks. Therefore, how to effectively exploit complementary information from such multi-view data remains a central problem in various applications. In recent years, many multi-view learning algorithms have been proposed to exploit the related structure of multi-view data to improve the performance of the learning tasks, where they assume that data from different views are corrected and aim to learn a shared latent space to discover such correlations.

Canonical correlation analysis (CCA) is one of the most representative learning method, which learn a couple of projections of two view to maximize their correlations in the latent space. While encouraging performance can be obtained, CCA only works well for two views and cannot be applicable to multiple views directly. To address this, Fu et al. [32] proposed a general subspace learning approach to perform multiple feature fusion with the CCA criterion over multiple pairs of views, Lu et al. [33] presented a multi-view neighborhood repulsed metric learning approach to learn a latent feature space to map features from different views into the common space. However, these multi-view learning methods assume that samples in the training and testing sets are collected under the same condition such that they are independent and identically distributed, which are not suitable for cross-dataset facial expression recognition because discriminative information exploited in the source domain cannot be applied to target domain directly. In this work, we develop a multi-view transfer subspace learning approach where multiple different yet related subspaces are learned to transfer information from the source domain to the target domain, so that complementary information for multiple feature representations of face images can be better exploited for facial expression recognition.

3. Transfer subspace learning

3.1. Basic idea

Conventional subspace learning algorithms seek a feature subspace W by solving an optimization objective function $F(W)$ and then apply W for feature extraction since the training and testing samples are implicitly assumed to be independent and identical distribution. As mentioned before, this assumption will not hold for cross-dataset facial recognition problem. Hence, we also need to minimize the difference between the training and testing sets besides optimizing $F(W)$.

Given N_1 training samples $X=[x_1,x_2,\dots,x_{N_1}]$ and N_2 testing samples $Y=[y_1,y_2,\dots,y_{N_2}]$, our objective now is seeking a feature space W to optimize $F(W)$ in the training set and minimize the differences between X and Y in W simultaneously. Specifically, we formulate our objective into the following optimization function:

$$\min_W J(W)=F(W)+\lambda H(W) \tag{3}$$

where $\lambda \ge 0$, and

$$H(W)=\sum_{i=1}^{N_1} \| W^T x_i - W^T \sum_{j=1}^{k} t_{ij} y_{ij} \|^2 \tag{4}$$

$y_{i1}, y_{i2}, \dots, y_{ik}$ are the k-nearest neighbors of x_i, $t_{i1}, t_{i2}, \dots, t_{ik}$ are the corresponding coefficients, and they can be obtained similarly to the coefficients obtained in the locally linear embedding (LLE) method in [24].

We simplify $H(W)$ to the following form:

$$\begin{aligned} H(W) &= \sum_{i=1}^{N_1} \mathrm{tr}\left[W^T\left(x_i-\sum_{j=1}^{k}t_{ij}y_{ij}\right)\left(x_i-\sum_{j=1}^{k}t_{ij}y_{ij}\right)^T W\right] \\ &= \mathrm{tr}\left[W^T\sum_{i=1}^{N_1}\left(x_i-\sum_{j=1}^{k}t_{ij}y_{ij}\right)\left(x_i-\sum_{j=1}^{k}t_{ij}y_{ij}\right)^T W\right] \\ &= \mathrm{tr}(W^T G W) \end{aligned} \tag{5}$$

where $G \triangleq \sum_{i=1}^{N_1}\left[\left(x_i-\sum_{j=1}^{k}t_{ij}y_{ij}\right)\left(x_i-\sum_{j=1}^{k}t_{ij}y_{ij}\right)^T\right]$.

The derivative of $\frac{\partial H(W)}{\partial W}$ is

$$\frac{\partial H(W)}{\partial W}=2GW \tag{6}$$

As different subspace learning methods have different $F(W)$, we include different $F(W)$ for different subspace learning methods and formulate the corresponding transferred ones in the following.

3.2. TPCA

From Table 1, we can obtain $F(W)=-\mathrm{tr}(W^T S_T W)$ for PCA. To make the minimization problem with respect to W well-posed, we impose an orthogonal constraint $W^T W=I$ and formulate TPCA as the following constrained optimization problem:

$$\begin{aligned} \min_W \quad & T(W)=-\mathrm{tr}(W^T S_T W)+\lambda\,\mathrm{tr}(W^T G W) \\ \text{s.t.} \quad & W^T W = I. \end{aligned} \tag{7}$$

Let $\frac{\partial T(W)}{\partial W}=0$, we can obtain the projections of TPCA by solving the following eigenvalue equation:

$$(\lambda G - S_T)w=\alpha w \tag{8}$$

Let $\{w_1,w_2,\dots,w_p\}$ be the eigenvectors corresponding to the p smallest eigenvalues $\{\alpha_i \mid i=1,2,\dots,k\}$ ordered such that $\alpha_1 \le \alpha_2 \le \dots \le \alpha_p$. Then $W=[w_1,w_2,\dots,w_p]$ is the subspace projection of TPCA.

3.3. TLDA

From Table 1, we can obtain $F(W)=\frac{\mathrm{tr}(W^T S_W W)}{\mathrm{tr}(W^T S_B W)}$ for LDA. Hence,

$$\frac{\partial F(W)}{\partial W}=2p_1^{-1}S_W W-2p_1^{-2}p_2 S_B W \tag{9}$$

where $p_1=\mathrm{tr}(W^T S_B W)$ and $p_2=\mathrm{tr}(W^T S_W W)$.

As Eq. (3) is nonlinear and it is nontrivial to derive its closed-form global optimal solution, we modified the trace ratio of LDA to the difference form and seek a global solution by the following optimization problem:

$$\begin{aligned} \min_W T(W) \quad & =\mathrm{tr}(W^T(S_W-S_B)W)+\lambda\,\mathrm{tr}(W^T G W) \\ \text{s.t.} \quad & W^T W = I. \end{aligned} \tag{10}$$

Let $\frac{\partial T(W)}{\partial W}=0$, we can obtain the projections of TLDA by solving the following eigenvalue equation

$$(\lambda G+S_W-S_B)w=\alpha w \tag{11}$$

We can obtain the projections of TLDA similarly to that of TPCA.

3.4. TLPP

For LPP, $F(W)=W^TXLX^TW$. Hence, TLPP can be formulated as the following constrained optimization problem:

$$\min_W \quad T(W)=W^TXLX^TW+\lambda\,\mathrm{tr}(W^TGW)$$
$$\text{s.t.} \quad W^TW=I. \tag{12}$$

Let $\frac{\partial T(W)}{\partial W}=0$, we can obtain the projections of TLPP by solving the following eigenvalue equation:

$$(XLX^T+\lambda G)w=\alpha w \tag{13}$$

We can obtain the projections of TLPP similarly to that of TPCA.

3.5. TONPP

For ONPP, $F(W)=tr(W^TXMX^TW)$. Hence, TONPP can be formulated as the following constrained optimization problem:

$$\min_W T(W) \quad =\mathrm{tr}(W^TXMX^TW)+\lambda\,\mathrm{tr}(W^TGW)$$
$$\text{s.t.} \quad W^TW=I. \tag{14}$$

Let $\frac{\partial T(W)}{\partial W}=0$, we can obtain the projections of TONPP by solving the following eigenvalue equation:

$$(XMX^T+\lambda G)w=\alpha w \tag{15}$$

We can obtain the projections of TONPP similarly to that of TPCA.

4. Multi-view transfer subspace learning

Assume there are K feature descriptors extracted for each face sample, $X^k=\{x_i^k|i=1,2,\ldots,N_1\}$ and $Y^k=\{y_i^k|i=1,2,\ldots,N_2\}$ be the training and testing samples in the kth view, the objective of our multi-view transfer subspace learning aims to learn K feature spaces $W_1,\ldots,W_K$ and a nonnegative weighting vector $\eta=[\eta_1,\eta_2,\ldots,\eta_K]$ to optimize $F(W_1,\ldots,W_K)$ in the training set and minimize the differences between X and Y in the learned feature spaces $W_1,\ldots,W_K$, which are defined as follows:

$$\min J(W_1,\ldots,W_K,\eta) \quad =Z(W_1,\ldots,W_K)+\delta R(W_1,\ldots,W_K)$$
$$\text{subject to} \quad \sum_{k=1}^{K}\eta_k=1, \quad \eta_k\ge 0, 1\le k\le K. \tag{16}$$

where $\delta\ge 0$, and

$$Z(W_1,\ldots,W_K)=\sum_{k=1}^{K}\eta_k(F_k(W_1,\ldots,W_K)+\lambda H_k(W_1,\ldots,W_K)) \tag{17}$$

$$R(W_1,\ldots,W_K)=\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{K}\sum_{i=1}^{N}\|W_{k_1}^Tx_i^{k_1}-W_{k_2}^Tx_i^{k_2}\|_2^2 \tag{18}$$

The physical meaning of (16) is to learn K subspaces W_k $(k=1,2,\ldots,K)$ under which:

(1) The difference of features of the same sample is enforced to be as small as possible, which is equivalent to the canonical correlation analysis based multiple feature fusion method where the correlation of different feature representations of each sample are maximized;

(2) Feature subspaces learned from the source domain can be transferred into the target domain.

The trivial solution of (16) is $\eta_k=1$, which corresponds to the minimum $Z(W_1,\ldots,W_K)$, and $\eta_k=0$ otherwise. This means that only the best single data is selected for recognition, To address this, we modify η_k to be $\eta_k{}^p$ $(p>1)$, and rewrite the following objective function:

$$\min_{W_1,\ldots,W_K,\eta} \quad \sum_{k=1}^{K}\eta_k^p(F_k(W_1,\ldots,W_K)+\lambda H_k(W_1,\ldots,W_K))$$
$$+\lambda\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{K}\sum_{i=1}^{N}\|W_{k_1}^Tx_i^{k_1}-W_{k_2}^Tx_i^{k_2}\|_2^2$$
$$\text{subject to} \quad \sum_{k=1}^{K}\eta_k=1,\eta_k\ge 0, \quad 1\le k\le K. \tag{19}$$

Since there are K matrices to be optimized simultaneously in (19), we use the following alternating optimization method to obtain a local solution.

First, we fix η and update W. When η is fixed, the optimization problem in (19) can be rewritten as:

$$\min_{W_1,W_2,\ldots,W_K}\sum_{k=1}^{K}\eta_k^p(F_k(W_1,\ldots,W_K)+\lambda H_k(W_1,\ldots,W_K))$$
$$+\lambda\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{K}\sum_{i=1}^{N}\|W_{k_1}^Tx_i^{k_1}-W_{k_2}^Tx_i^{k_2}\|_2^2 \tag{20}$$

We sequentially optimize W_k with the fixed W_1, W_2, ⋯, W_{k-1}, W_{k+1}, ⋯, W_K. Then, (20) can be rewritten as:

$$\min_{W_k} J(W_k)=\eta_k^p(F_k(W_k)+\lambda H_k(W_k))+\lambda\sum_{\substack{k_1,k=1\\k_1\neq k}}^{K}\sum_{i=1}^{N}\|W_k^Tx_i^k-W_{k_1}^Tx_i^{k_1}\|_2^2 \tag{21}$$

W_k can be obtained by solving the following gradient descent algorithm:

$$W_k=W_k-\theta\frac{\partial J(W_k)}{\partial W_k} \tag{22}$$

where $\frac{\partial J(W_k)}{\partial W_k}$ is the gradient of the objective function $J(W_k)$ with respect to the parameter W_k, and θ is the learning rate.

Then, we update η with the fixed W_k. We construct the following Lagrange function:

$$J(\eta,\zeta)=\sum_{k=1}^{K}\eta_k^pg_k(W_k)+\lambda\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{K}\sum_{i=1}^{N}\|W_{k_1}^Tx_i^{k_1}-W_{k_2}^Tx_i^{k_2}\|_2^2$$
$$-\zeta\left(\sum_{k=1}^{K}\eta_k-1\right) \tag{23}$$

where $g_k=F_k(W_k)+\lambda H_k(W_k)$.

Let $\frac{\partial J(\eta,\zeta)}{\partial\eta_k}=0$ and $\frac{\partial J(\eta,\zeta)}{\partial\zeta}=0$, we have

$$p\eta_k^{p-1}g_k(W_k)-\zeta=0 \tag{24}$$

$$\sum_{k=1}^{K}\eta_k-1=0 \tag{25}$$

Combining (24) and (25), we can obtain η_k as follows

$$\eta_k=\frac{(1/g_k(W_k))^{1/(p-1)}}{\sum_{k=1}^{K}(1/g_k(W_k))^{1/(p-1)}}. \tag{26}$$

We repeat the above procedure until the algorithm converges. The proposed Multi-view Transfer Subspace Learning algorithm is summarized in **Algorithm 2**, where $E^{d_l \times d_k}$ is a matrix with ones on the diagonal and zeros elsewhere.

Algorithm 1. Multi-view Transfer Subspace Learning.

Input: Training set $X = \{X^k\}_{k=1}^{K}, X^k \in R^{d_k \times N}$ is the kth feature set, learning rate θ, parameter p and λ, and convergence error τ.
Output: Subspaces: $W_1, W_2, \ldots, W_K$ and weights: $\eta_1, \eta_2, \ldots, \eta_K$.
Step 1 (Initialization):
1.1. Set $\eta_k = 1/K$, $L_k = E^{d_l \times d_k}$, $k = 1, \ldots, K$.
Step 2 (Local optimization):
For $r = 1, 2, \ldots,$ repeat
2.1. Compute W_k by using (WW).
2.2. Obtain η_k by using (26).
2.3. If $r > 2$ and $|W^r - W^{r-1}| < \tau$, go to Step 3.
Step 3 (Output subspaces and weighting vector):
Output $W_1, W_2, \ldots, W_K$ and $\eta_1, \eta_2, \ldots, \eta_K$.

5. Experimental results

5.1. Datasets and settings

Three publicly available facial expression image databases including the JAFFE [25,26], Cohn–Kanade [27], Feedtum [28] databases were selected to evaluate the effectiveness of the proposed methods for cross-dataset facial expression recognition.

The JAFFE database consists of 213 facial expression images from 10 Japanese females. They posed 3 or 4 examples for each of the seven basic expressions (six emotional expressions including anger, disgust, fear, happy, sad, surprise plus neutral expression). The image size is 256×256.

The Cohn–Kanade database consists of 100 university students aged from 18 to 30 years. 65% subjects are female, 15% are African-American, and 3% are Asian or Latino. Subjects are instructed to perform a series of 23 facial displays, seven of which are anger, disgust, fear, happy, neural, sad and surprise. We selected 10 subjects which contain all the seven different expressions from the database, where each expression has four samples. Hence, we have 280 samples in total. As the original image sequences in the database start from a neural expression and end with the peak of the expression, we selected the last four frames of each expression sequence. For the neural expression, we selected the first frame of four different sequences. The size of the original facial image is 640×490.

The Feedtum database, also known as the FG-NET database, is much more challenging because in the database subjects performed the expressions spontaneously and some of the resulting expressions are not well distinguishable. It contains a set of facial image sequences that show a number of subjects performing the seven different universal expressions defined by Ekman and Friesen. All seven expressions were performed three times by each subject. Since these images were captured under natural circumstances, there could be head movement in the images. However, in order to simplify our experiments, only the images which include frontal faces without large head movement were chosen. We selected 10 subjects which contain all the seven different expressions from the database, where each expression has four samples. Hence, we have 280 samples in total. The size of the original facial image is 320×240.

For all the three databases, we converted the images to gray scale and manually located the eye positions. We cropped the face regions from original images according to the eyes' positions and resized them to 64×64. No further registration such as alignment of mouth was performed in our experiments. Some examples of the aligned images from the databases are shown in Fig. 1, where 1(a), 1(b) and 1(c) are the example samples of the JAFFE, Cohn–Kanade and Feedtum databases, respectively.

For transfer subspace learning methods, the raw pixels are used for face representation. For multi-view transfer subspace learning, we apply three more different feature descriptors including Local Binary Patterns (LBP) [34], Spatial Pyramid LEarning (SPLE) [35] and Scale-Invariant Feature Transform (SIFT) [36] to extract different and complementary information from each face image. The reason we selected these three features is that they have shown reasonably good performance in recent facial expression recognition studies. No doubt, more effective feature descriptors could be employed to improve the verification performance. However, the main interest in this study is to evaluate the proposed multi-view transfer subspace learning methods which use multiple features for facial expression recognition.

For each face image, we employed 256 bins to extract the LBP feature. For the SPLE feature, three different resolutions are first

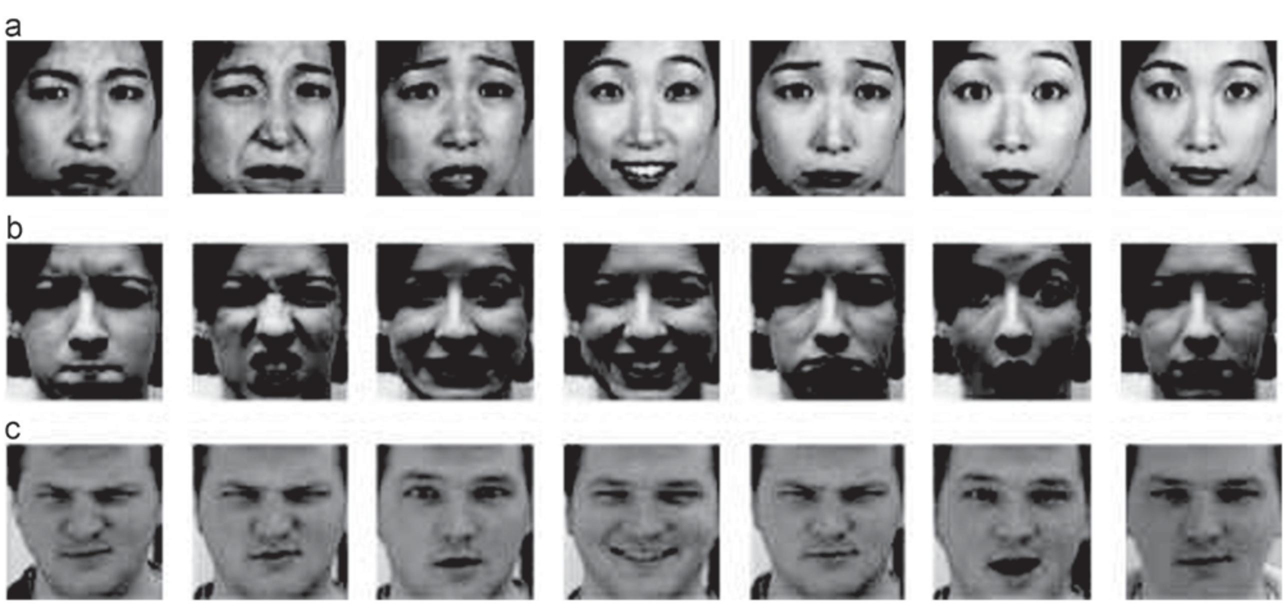


Fig. 1. Facial expression images of one subject from (a) JAFFE, (b) Cohn–Kanade, and (c) Feedtum databases. From left to right are the images with anger, disgust, fear, happy, sad, surprise and neural expressions, respectively.

Please cite this article as: H. Yan, Transfer subspace learning for cross-dataset facial expression recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.11.113

constructed and 21 cells are obtained. Then, each local feature in each cell was quantized into 200 bins and each face image was represented by a 4200-dimensional long feature vector. For the SIFT feature, each SIFT descriptors was first sampled over each 16×16 patch with a grid spacing of 8 pixels. Then, each SIFT descriptor is concatenated into a long feature vector. For these features, we apply PCA to reduce each feature into 200 dimensions to remove some noise components.

We employed the nearest neighbor (NN) classifier for facial expression recognition. The value of λ was empirically set to be 10 for all the four transfer subspace learning methods. We compared our proposed transfer subspace learning methods with four existing non-transferred subspace learning methods including PCA, LDA, LPP and ONPP for cross-dataset facial expression recognition. The feature dimensions of all subspace learning methods are set as 200 except the LDA related methods, where the feature dimensions of these methods are set to 6.

Based on the three datasets, we conducted six sets of cross-dataset facial expression recognition as follows:

1. J2C: the training set is JAFFE and the testing set is Cohn–Kanade;
2. J2F: the training set is JAFFE and the testing set is Feedtum;
3. C2J: the training set is Cohn–Kanade and the testing set is JAFFE;
4. C2F: the training set is Cohn–Kanade and the testing set is Feedtum;
5. F2J: the training set is Feedtum and the testing set is JAFFE;
6. F2C: the training set is Feedtum and the testing set is Cohn–Kanade.

5.2. Results

The confusion matrices of the seven expressions under the F2C setting were also calculated for PCA, LDA, LPP, ONPP, TPCA, TLDA, TLPP, TONPP, MTPCA, MTLDA, MTLPP, and MTONPP and tabulated in Tables 2–13, respectively, where ANG, DIS, FEA, HAP, SAD, SUR and NEU represents the anger, disgust, fear, happy, sad, surprise and neural expressions, respectively. We can observe from these results that diagonal elements of the confusion matrices of transfer subspace learning methods are generally better than those of conventional non-transferred subspace learning methods, which further indicated that transfer subspace learning approach can improve the recognition accuracy of subspace learning for cross-dataset facial expression recognition. That is because conventional subspace learning algorithms such as PCA, LDA, LPP and ONPP assume that the training and testing samples are independent and identically distributed variables and this assumption does not hold for cross-dataset facial expression recognition tasks. Moreover, multi-view transfer subspace learning methods achieve higher performance than the corresponding single-view transfer subspace learning methods.

Table 2
Confusion matrix of seven-class expression recognition obtained by PCA under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**30.3%**	22.3%	6.4%	1.4%	20.3%	7.3%	12.0%
DIS	3.6%	**29.8%**	22.4%	1.8%	22.6%	14.3%	5.5%
FEA	13.6%	21.4%	**27.6%**	21.8%	5.2%	7.4%	3.0%
HAP	3.0%	13.6%	21.4%	**28.6%**	20.8%	8.2%	4.4%
SAD	8.6%	16.2%	18.8%	5.6%	**29.6%**	15.2%	6.0%
SUR	3.0%	13.2%	21.8%	18.8%	9.2%	**28.6%**	5.4%
NEU	6.4%	15.3%	12.4%	11.4%	10.2%	19.3%	**25.0%**

Table 3
Confusion matrix of seven-class expression recognition obtained by LDA under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**37.0%**	20.3%	6.4%	1.4%	16.3%	7.3%	11.3%
DIS	3.6%	**36.5%**	20.4%	1.8%	20.6%	12.3%	4.8%
FEA	10.6%	20.4%	**34.3%**	19.8%	5.2%	7.2%	2.5%
HAP	3.0%	11.6%	19.4%	**35.3%**	18.8%	7.7%	4.2%
SAD	6.6%	15.2%	16.8%	5.6%	**36.3%**	14.2%	5.3%
SUR	3.0%	12.2%	19.8%	16.8%	8.2%	**35.3%**	4.7%
NEU	5.4%	13.3%	10.4%	11.4%	8.5%	17.3%	**33.7%**

Table 4
Confusion matrix of seven-class expression recognition obtained by LPP under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**34.0%**	20.3%	6.4%	1.4%	19.3%	7.3%	11.3%
DIS	3.5%	**33.5%**	20.4%	1.8%	21.6%	14.2%	5.0%
FEA	12.6%	20.4%	**31.3%**	20.8%	5.0%	6.9%	3.0%
HAP	3.0%	12.6%	20.4%	**32.3%**	19.1%	8.2%	4.4%
SAD	7.6%	15.2%	17.8%	5.3%	**33.3%**	14.8%	6.0%
SUR	3.0%	12.2%	20.8%	17.8%	8.5%	**32.3%**	5.4%
NEU	5.4%	15.3%	11.4%	9.4%	11.2%	18.6%	**28.7%**

Table 5
Confusion matrix of seven-class expression recognition obtained by ONPP under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**39.6%**	17.3%	6.4%	1.4%	16.3%	7.2%	11.8%
DIS	3.6%	**39.1%**	18.4%	1.2%	20.6%	12.3%	4.8%
FEA	10.6%	18.4%	**36.9%**	19.2%	5.2%	7.2%	2.5%
HAP	3.0%	11.6%	18.4%	**37.9%**	17.2%	7.7%	4.2%
SAD	6.6%	15.2%	15.8%	5.6%	**38.9%**	13.2%	4.7%
SUR	3.0%	12.2%	17.8%	16.2%	8.2%	**37.9%**	4.7%
NEU	5.4%	13.3%	10.4%	10.8%	8.5%	15.3%	**36.3%**

Table 6
Confusion matrix of seven-class expression recognition obtained by TPCA under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**45.3%**	12.3%	6.4%	1.4%	15.3%	7.3%	12.0%
DIS	3.6%	**44.8%**	12.4%	1.8%	17.6%	14.3%	5.5%
FEA	13.6%	11.4%	**42.6%**	16.8%	5.2%	7.4%	3.0%
HAP	3.0%	13.6%	11.4%	**43.6%**	15.8%	8.2%	4.4%
SAD	8.6%	11.2%	13.8%	5.6%	**44.6%**	10.2%	6.0%
SUR	3.0%	13.2%	11.8%	13.8%	9.2%	**43.6%**	5.4%
NEU	6.4%	10.3%	12.4%	10.4%	10.2%	10.3%	**40.0%**

Table 7
Confusion matrix of seven-class expression recognition obtained by TLDA under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**55.0%**	10.3%	6.4%	1.4%	10.3%	7.5%	9.1%
DIS	3.6%	**54.5%**	10.4%	1.8%	12.6%	12.3%	4.8%
FEA	10.6%	10.4%	**52.3%**	12.8%	5.2%	6.2%	2.5%
HAP	3.0%	10.6%	10.4%	**53.3%**	10.8%	7.7%	4.2%
SAD	3.6%	10.2%	10.8%	5.6%	**54.3%**	10.2%	5.3%
SUR	3.0%	12.2%	9.8%	8.8%	8.2%	**53.3%**	4.7%
NEU	5.4%	10.3%	6.4%	10.4%	8.5%	7.3%	**51.7%**

Please cite this article as: H. Yan, Transfer subspace learning for cross-dataset facial expression recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.11.113

Table 8
Confusion matrix of seven-class expression recognition obtained by TLPP under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**51.0%**	10.3%	6.4%	1.4%	12.3%	7.3%	11.3%
DIS	3.6%	**50.5%**	10.4%	1.8%	14.6%	14.1%	5.0%
FEA	12.6%	10.4%	**48.3%**	13.8%	5.0%	6.9%	3.0%
HAP	3.0%	12.6%	10.4%	**49.3%**	12.1%	8.2%	4.4%
SAD	7.6%	10.2%	10.8%	5.3%	**50.3%**	9.8%	6.0%
SUR	3.0%	12.2%	10.8%	10.8%	8.5%	**49.3%**	5.4%
NEU	5.4%	1.3%	9.4%	9.4%	10.2%	8.6%	**55.7%**

Table 9
Confusion matrix of seven-class expression recognition obtained by TONPP under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**55.6%**	9.3%	6.4%	1.4%	8.3%	7.2%	11.8%
DIS	3.6%	**55.1%**	11.4%	1.2%	11.6%	12.3%	4.8%
FEA	10.6%	10.4%	**52.9%**	11.2%	5.2%	7.2%	2.5%
HAP	3.0%	11.6%	11.4%	**53.9%**	8.2%	7.7%	4.2%
SAD	6.6%	10.2%	11.8%	5.6%	**54.9%**	8.2%	2.7%
SUR	3.0%	7.2%	11.8%	11.2%	8.2%	**53.9%**	4.7%
NEU	5.4%	11.3%	6.4%	6.8%	8.5%	9.3%	**52.3%**

Table 10
Confusion matrix of seven-class expression recognition obtained by MTPCA under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**48.3%**	11.3%	5.4%	1.4%	14.3%	7.3%	12.0%
DIS	3.6%	**46.8%**	11.4%	1.8%	11.6%	14.3%	5.5%
FEA	11.6%	11.4%	**45.6%**	14.8%	5.2%	7.4%	3.0%
HAP	3.0%	11.6%	11.4%	**47.6%**	13.8%	8.2%	4.4%
SAD	8.6%	10.2%	11.8%	5.6%	**47.6%**	10.2%	6.0%
SUR	3.0%	11.2%	10.8%	13.8%	9.2%	**46.6%**	5.4%
NEU	6.4%	9.3%	10.4%	10.4%	10.2%	10.3%	**43.0%**

Table 11
Confusion matrix of seven-class expression recognition obtained by MTLDA under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**58.0%**	9.3%	4.4%	1.4%	10.3%	7.5%	9.1%
DIS	3.6%	**57.5%**	10.4%	1.8%	12.6%	10.3%	3.8%
FEA	10.6%	9.4%	**55.3%**	10.8%	5.2%	6.2%	2.5%
HAP	3.0%	10.6%	9.4%	**55.3%**	9.8%	7.7%	4.2%
SAD	3.6%	9.2%	10.8%	5.6%	**57.3%**	8.2%	5.3%
SUR	3.0%	11.2%	7.8%	8.8%	8.2%	**56.3%**	4.7%
NEU	5.4%	9.3%	6.4%	8.4%	8.5%	7.3%	**54.7%**

Fig. 2 shows the mean verification rate of TPCA and MTPCA methods versus different number of iterations under the F2C setting. Our methods obtain stable recognition rates in several iterations.

We also examine the mean recognition rate of our multi-view transfer subspace learning methods versus different values of p under the F2C setting. We find that the performance of our multi-view transfer subspace learning methods are not sensitive to this parameter and the best performance can be obtained when p was set to 3.

Table 12
Confusion matrix of seven-class expression recognition obtained by MTLPP under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**55.0%**	8.3%	6.4%	1.4%	12.3%	7.3%	9.3%
DIS	3.6%	**54.5%**	8.4%	1.8%	14.6%	12.1%	5.0%
FEA	11.6%	10.4%	**51.3%**	11.8%	5.0%	6.9%	3.0%
HAP	3.0%	11.6%	10.4%	**52.3%**	10.1%	8.2%	4.4%
SAD	7.6%	9.2%	9.8%	5.3%	**53.3%**	8.8%	6.0%
SUR	3.0%	11.2%	9.8%	9.8%	8.5%	**52.3%**	5.4%
NEU	5.4%	1.3%	8.4%	8.4%	9.2%	8.6%	**58.7%**

Table 13
Confusion matrix of seven-class expression recognition obtained by MTONPP under the F2C setting.

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	**58.6%**	9.3%	6.4%	1.4%	7.3%	6.2%	10.8%
DIS	3.6%	**58.1%**	10.4%	1.2%	10.6%	11.3%	4.8%
FEA	9.6%	9.4%	**55.9%**	10.2%	5.2%	7.2%	2.5%
HAP	3.0%	10.6%	10.4%	**55.9%**	8.2%	7.7%	4.2%
SAD	6.6%	8.2%	10.8%	5.6%	**57.9%**	8.2%	2.7%
SUR	3.0%	7.2%	10.8%	9.2%	8.2%	**56.9%**	4.7%
NEU	5.4%	10.3%	5.4%	5.8%	7.5%	9.3%	**56.3%**

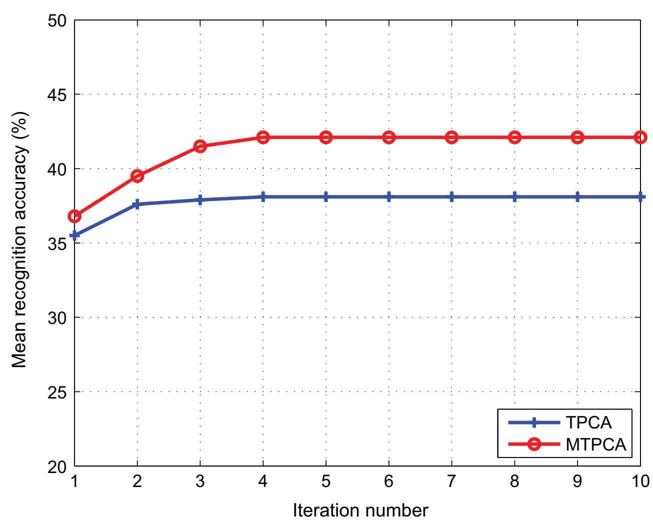


Fig. 2. Mean recognition rates (%) our transfer subspace learning approach versus different number of iterations under the F2C setting.

6. Conclusions and future work

We have investigated in this paper the problem of cross-dataset facial expression recognition. Since the training and testing samples are not independent and identically distributed in many real facial expression recognition applications, we have proposed a new transfer subspace learning approach to learn a feature space which transfers the knowledge gained from the training set to the target (testing) data to improve the recognition performance under cross-dataset scenarios. Following this idea, we have formulated four new transfer subspace learning methods, i.e.,

Please cite this article as: H. Yan, Transfer subspace learning for cross-dataset facial expression recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.11.113

transfer PCA (TPCA), transfer LDA (TLDA), transfer LPP (TLPP), and transfer ONPP (TONPP) for cross-dataset facial expression recognition. To better exploit more complementary information for multiple feature representations of face images, we have developed a multi-view transfer subspace learning approach where multiple different yet related subspaces are learned to transfer information from the source domain to the target domain, where four multi-view transfer subspace learning methods, namely multi-view transfer PCA (MTPCA), multi-view transfer LDA (MTLDA), multi-view transfer LPP (MTLPP), and multi-view transfer ONPP (MTONPP) are introduced. Experimental results have demonstrated the efficacy of the proposed methods.

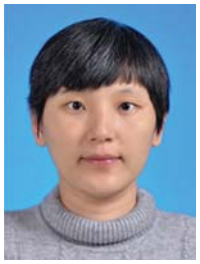
1. In this work, a linear model is employed to transfer information exploited in the source domain to the target domain. For future work, we are interested in using a nonlinear model for transfer learning.
2. In this work, a linear combination of multi-view data is employed in our multi-view transfer subspace learning approach. It is interesting to apply a nonlinear combination model to better exploit the complementary information of different views.
3. It is interesting to apply our proposed methods to cross-dataset visual recognition applications such as cross-domain person re-identification and object recognition to further demonstrate their effectiveness.

References

[1] M. Pantic, L.J.M. Rothkrantz, Automatic analysis of facial expressions: the state of the art, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1424–1445.
[2] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern Recognit. 36 (2003) 259–275.
[3] Z. Zeng, M. Pantic, G.I. Roisman, T. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Trans. Pattern Anal. Mach. Intell. 31 (1) (2009) 39–58.
[4] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fwllenz, J. G. Taylor, Emotion recognition in human-computer interaction, IEEE Signal Process. Mag. 18 (1) (2001) 32–80.
[5] I. Kotsia, I. Pitas, Facial expression recognition in image sequences using geometric deformation features and support vector machines, IEEE Trans. Image Process. 16 (1) (2007) 160–187.
[6] Y. Gao, M.K.H. Leung, S.C. Hui, M.W. Tananda, Facial expression recognition from line-based caricatures, IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 33 (3) (2003) 407–412.
[7] Y. Cheon, D. Kim, Natural facial expression recognition using differential-AAM and manifold learning, Pattern Recognit. 42 (2009) 1340–1350.
[8] B. Raducanu, F. Dornaika, Dynamic facial expression recognition using Laplacian eigenmaps-based manifold learning, Proceedings of IEEE International Conference on Robotics and Automation, 2010, pp. 156–161.
[9] C. Shan, S. Gong, P.W. McOwan, A comprehensive empirical study on linear subspace methods for facial expression analysis, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshop, 2006, pp. 153–158.
[10] C. Shan, S. Gong, P.W. McOwan, Appearance manifold of facial expression, Lecture Notes in Computer Science 3766 (2005) 221–230.
[11] S. Zafeiriou, I. Pitas, Discriminant graph structures for facial expression recognition, IEEE Trans. Multimedia 10 (8) (2008) 1528–1540.
[12] R. Zhi, Q. Ruan, Discriminant spectral analysis for facial expression recognition, in: Proceedings of IEEE International Conference on Image Processing, 2008, pp. 1924–1927.
[13] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image Vis. Comput. 27 (2009) 803–816.
[14] G. Zhao, M. Pietikainen, Boosted multi-resolution spatiotemporal descriptors for facial expression recognition, Pattern Recognit. Lett. 30 (2009) 1117–1127.
[15] W. Zheng, X. Zhou, C. Zou, L. Zhao, Facial expression recognition using kernel canonical correlation analysis (KCCA), IEEE Trans. Neural Netw. 17 (1) (2006) 233–238.
[16] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1) (1991) 71–86.
[17] P.N. Belhumenur, J.P. Hepanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[18] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, Face recognition using laplacian faces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[19] E. Kokiopoulou, Y. Saad, Orthogonal neighborhood preserving projections, in: Proceedings of International Conference on Data Mining, 2005, pp. 234–241.
[20] R. Xiao, Q. Zhao, D. Zhang, P. Shi, Facial expression recognition on multiple manifolds, Pattern Recognit. 44 (1) (2011) 107–116.
[21] J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359.
[22] S. Si, D. Tao, B. Geng, Bregman divergence-based regularization for transfer subspace learning, IEEE Trans. Knowl. Data Eng. 22 (7) (2010) 929–942.
[23] Y. Su, Y. Fu, Q. Tian, X. Gao, Cross-database age estimation based on transfer learning, Proceedings of International Conference on Acoustics, Speech and Signal Processing, 2010, pp. 1270–1273.
[24] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[25] M.J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with Gabor wavelets, in: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 200–220.
[26] M.J. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Trans. Pattern Anal. Mach. Intell. 21 (12) (1999) 1357–1362.
[27] T. Kanade, J. Cohn, Y.L. Tian, Comprehensive database for facial expression analysis, in: Proceedings of IEEE International Conference on Face and Gesture Analysis, 2000, pp. 46–53.
[28] F. Wallhoff, Facial expressions and memotion database, 〈Available at http://www.mmk.ei.tum.de/waf/fgnet/feedtum.html〉.
[29] J. Lu, Y.-P. Tan, A doubly weighted approach for appearance-based subspace learning methods, IEEE Trans. Inf. Forensics Secur. 5 (1) (2010) 71–81.
[30] J. Lu, Y.-P. Tan, Regularized locality preserving projections and its extensions for face recognition, IEEE Trans. Syst. Man Cybern.-Part B Cybern. 40 (3) (2010) 958–963.
[31] H. Yan, M.H. Ang, A.N. Poo, Cross-dataset facial expression recognition, in: IEEE International Conference on Robotics and Automation, 2011, pp. 5985–5990.
[32] Y. Fu, L. Cao, G. Guo, T.S. Huang, Multiple feature fusion by subspace learning, in: ACM International Conference on Image and Video Retrieval, 2008, pp. 127–134.
[33] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, J. Zhou, Neighborhood repulsed metric learning for kinship verification, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 331–345.
[34] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[35] X. Zhou, J. Hu, J. Lu, Y. Shang, Y. Guan, Kinship verification from facial images under uncontrolled conditions, in: ACM International Conference on Multimedia, 2011, pp. 953–956.
[36] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91110.
[37] J. Lu, V.E. Liong, X. Zhou, J. Zhou, Learning compact binary face descriptor for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2041–2256.
[38] J. Lu, Y.P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 39–51.
[39] J. Lu, G. Wang, P. Moulin, Human identity and gender recognition from gait sequences with arbitrary walking directions, IEEE Trans. Inf. Forensics Secur. 9 (1) (2014) 41–51.
[40] J. Lu, Y.P. Tan, Cost-sensitive subspace analysis and extensions for face recognition, IEEE Trans. Inf. Forensics Secur. 8 (3) (2013) 510–519.
[41] Y. Yan, E. Ricci, G. Liu, N. Sebe, Egocentric daily activity recognition via multitask clustering, IEEE Trans. Image Process. 24 (10) (2015) 2984–2995.
[42] Y. Yan, Y. Yang, D. Meng, G. Liu, W. Tong, A. Hauptmann, N. Sebe, Event oriented dictionary learning for complex event detection, IEEE Trans. Image Process. 24 (6) (2015) 1867–1878.
[43] Y. Yan, E. Ricci, R. Subramanian, G. Liu, N. Sebe, Multi-task linear discriminant analysis for multi-view action recognition, IEEE Trans. Image Process. 23 (12) (2014) 5599–5611.
[44] Y. Yan, E. Ricci, R. Subramanian, O. Lanz, N. Sebe, No matter where you are: flexible graph-guided multi-task learning for multi-view head pose classification under target motion, in: IEEE International Conference on Computer Vision, 2013.
[45] Y. Yan, H. Shen, G. Liu, Z. Ma, C. Gao, N. Sebe, Glocal tells you more: Coupling glocal structural for feature selection with sparsity for image and video classification, Comput. Vis. Image Understand. 125 (7) (2014) 99–109.
[46] Q. Zhao, D. Meng, Z. Xu, W. Zuo, Y. Yan, L1-Norm low-rank matrix factorization by variational bayes, IEEE Trans. Neural Netw. Learn. Syst. 26 (4) (2014) 825–839.
[47] J. Lu, V.E. Liong, J. Zhou, Cost-sensitive local binary feature learning for facial age estimation, IEEE Trans. Image Process. 24 (12) (2015) 5356–5368.
[48] J. Lu, G. Wang, P. Moulin, Localized multi-feature metric learning for image set based face recognition, IEEE Trans. Circuits Syst. Video Technol. (2015), accepted.
[49] J. Lu, G. Wang, W. Deng, K. Jia, Reconstruction-based metric learning for unconstrained face verification, IEEE Trans. Inf. Forensics Secur. 10 (1) (2015) 79–89.
[50] J. Lu, Y.-P. Tan, G. Wang, G. Yang, Image-to-set face recognition using locality repulsion projections and sparse reconstruction-based similarity measure, IEEE Trans. Circuits Syst. Video Technol. 23 (6) (2013) 1070–1080.
[51] J. Lu, Y.-P. Tan, Ordinary preserving manifold analysis for human age and head pose estimation, IEEE Trans. Hum.–Mach. Syst. 43 (2) (2013) 249–258.
[52] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, J. Zhou, Cost-sensitive semi-supervised discriminant analysis for face recognition, IEEE Trans. Inf. Forensics Secur. 7 (3) (2012) 944–953.
[53] J. Lu, V.E. Liong, G. Wang, P. Moulin, Joint feature learning for face recognition, IEEE Trans. Inf. Forensics Secur. 10 (7) (2015) 1371–1383.

Please cite this article as: H. Yan, Transfer subspace learning for cross-dataset facial expression recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.11.113

[54] H. Yan, J. Lu, X. Zhou, Prototype-based discriminative feature learning for kinship verification, IEEE Trans. Cybern. 45 (11) (2015) 2535–2545.
[55] H. Yan, J. Lu, W. Deng, X. Zhou, Discriminative multimetric learning for kinship verification, IEEE Trans. Inf. Forensics Secur. 9 (7) (2014) 1169–1178.
[56] J. Lu, Y.-P. Tan, Nearest feature space analysis for classification, IEEE Signal Process. Lett. 18 (1) (2011) 55–58.
[57] J. Lu, Y.-P. Tan, Uncorrelated discriminant nearest feature line analysis for face recognition, IEEE Signal Process. Lett. 17 (2) (2010) 185–188.
[58] J. Lu, E. Zhang, Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion, Pattern Recognit. Lett. 28 (16) (2007) 2401–2411.



Haibin Yan received the B.Eng. and M.Eng. degrees from the Xi'an University of Technology, Xi'an, China, in 2004 and 2007, and the Ph.D. degree from the National University of Singapore, Singapore, in 2013, all in mechanical engineering. Now, she is an assistant professor in the School of Automation, Beijing University of Posts and Telecommunications, Beijing, China. From October 2013 to July 2015, she was a research fellow at the Department of Mechanical Engineering, National University of Singapore, Singapore. Her research interests include social robotics and computer vision.

Please cite this article as: H. Yan, Transfer subspace learning for cross-dataset facial expression recognition, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2015.11.113