

Research Article

Subspace Learning and Joint Distribution Adaptation for Unsupervised Cross-Database Microexpression Recognition

Yanliang Zhang , Ying Liu, Geng Li, and Hongxing Peng 

School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454150, China

Correspondence should be addressed to Hongxing Peng; phx@hpu.edu.cn

Received 24 February 2021; Revised 6 May 2021; Accepted 25 June 2021; Published 20 July 2021

Academic Editor: Floriano Scioscia

Copyright © 2021 Yanliang Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Microexpression recognition has been widely favored by researchers due to its many potential applications, such as business negotiation and lie detection. Cross-database microexpression recognition is more challenging and attractive than normal microexpression recognition because the training and testing samples come from different databases. The ensuing challenge is that the feature distributions between training and testing samples differ too much. As a result, the performance of current well-performing microexpression recognition methods often fails to achieve the desired effect. In this paper, we overcome this problem by introducing Subspace Learning and Joint Distribution Adaptation (SLJDA) by projecting the source and target domains into the subspace and later reducing the distance between them and then minimizing the distance between the marginal and conditional probability distributions of the data between the source domain and the target domain. To evaluate its performance, a large number of cross-database experiments are performed in the SMIC database and CASMEII database. The experimental results show the superiority of the method compared with existing microexpression recognition methods.

1. Introduction

Microexpression is a special, weak facial expression that usually lasts only 1/25 to 1/5 seconds [1]. Different from ordinary facial expressions, microexpressions are spontaneous expressions that occur when people try to hide their inner emotions [2]. In other words, microexpressions can show the true thoughts of a person's heart. Therefore, automatic microexpression recognition technology can be applied in many practical scenarios such as marital relationship prediction [3], clinical diagnosis, and teaching assessment [4, 5], and even in the future, it may be extended to communication security [6–9], intelligent devices, etc.

In recent years, with the development of intelligent technology, researchers have gradually focused their attention on it, and microexpression recognition has made some progress. However, both training and testing samples for microexpression recognition come from the same microexpression database. This is not the case in practical applications, where the training and testing samples for

microexpression recognition come from different databases. Different database recording environments, camera devices, genders, ages, and so on will vary. They lead to the destruction of the consistency of feature distribution that exists in traditional microexpression recognition methods. Therefore, the microexpression recognition methods with good performance often fail to achieve the ideal effect in many practical applications. In order to solve this problem, it is necessary to study cross-database microexpression recognition; that is, the training set (train domain) and the testing set (target domain) are not the same database.

Yan et al. [10] classified cross-database facial expression recognition into two cases, semisupervised and unsupervised. Similarly, according to the presence or absence of label information in the target domain, cross-database microexpression recognition can also be divided into two categories, unsupervised cross-database microexpression recognition (CDMER) and semisupervised cross-database microexpression recognition. Unsupervised cross-database microexpression recognition can only be trained using the

sample label information of the source domain, while the latter can also incorporate a small amount of label information of the target domain. Therefore, this paper will focus on unsupervised cross-database microexpression recognition.

Domain adaptation is one of transfer learning methods [11], aiming at how to reduce the distribution differences between source and target domains data under the assumption that the source and target domain feature space and category space are the same and the feature distribution is different. In recent two years, Zong et al. [12] proposed cross-database microexpression recognition based on domain adaptive method for the first time and achieved good results. Current transfer learning methods can be roughly divided into three categories: data distribution [13–17], feature selection [18, 19], and subspace learning [20–23]. At present, the classical work based on these methods includes Transfer Component Analysis (TCA) [13], which finds a feature mapping while retaining the original information of the source domain and the target domain, so that the conditional distribution between the two domains after the mapping is relatively close. But such feature mapping may not exist. The feature selection-based approach, Structural Correspondence Learning (SCL) [19], assumes the existence of common data features in the source and target domains and uses shared features for modeling. However, when the characteristics of the two are very different, it can lead to failure. Based on the subspace learning method Subspace Alignment (SA) [20], the SA method is to directly seek a linear transformation by projecting the source and target domains into a subspace, through which the transformation alignment of different data is realized. However, such methods tend to ignore the distribution shift between the projected data in the two subspaces. In the face of this challenge, we integrate the subspace learning method with the data distribution adaptation method, namely, Subspace Learning and Joint Distribution Adaptation (SLJDA). First, the source and target domains are mapped to the corresponding subspaces, followed by maximizing the interclass distance and minimizing the intraclass distance in turn to preserve the discriminative features of the source domain data. Maximize the target domain variance samples, minimize the differences in the distribution of the corresponding subspaces of the projected source and target domains, and minimize the distances between subspaces to reduce the difference between source and target domains and improve the recognition of unsupervised cross-database microexpression.

In general, this paper contains the following main contributions:

- (1) After considering the practical situation, the subspace-based learning method and the data distribution-based adaptive method are cleverly combined to reduce the difference between the source and target domains. The original information is preserved, and there is no need to search for feature mappings that may not exist to a large extent. In addition, the data distribution shift in the projected

subspace is reduced, and the distance between them is geometrically reduced.

- (2) We conducted more extensive CDMER experiments to assess the performance of the SLJDA approach.

The rest of this paper is arranged as follows. Section 2 reviews the recent research on unsupervised cross-database microexpression. In Section 3, we describe the SLJDA for CDMER in detail. For better evaluating this method, extensive experiments and analyses on SMIC and CASMEII are shown in Section 4. Finally, the conclusion and the future planning of this paper are drawn in Section 5.

2. Related Work

As mentioned above, most of the existing cross-database emotion recognition problems (including cross-database voice emotion recognition, facial expression recognition, and microexpression recognition) usually solve the mismatch of feature distribution between source domain and target domain by domain adaptive DA method. In the recent two years, a domain adaptive approach was first proposed by Zong et al. [12]. In addition, Target Sample Re-Generator (TSRG) was proposed to learn a sample regenerator that can regenerate the samples of source domain and target domain, and the samples of both have the same characteristic distribution, where the regenerated source domain microexpression samples will remain unchanged in the feature space. Li et al. [24] proposed to use the Target-Adapted Least-Squares Regression (TALSR) method to learn the regression coefficient matrix according to the source domain data information and make the regression coefficient matrix fit the target database.

Before this, most researchers focused on cross-database facial expression and cross-database speech recognition related to cross-database microexpression and put forward many effective methods. For example, Zhu et al. [25] first proposed to use domain adaptation to solve the facial expression recognition problem and achieve good results. In the work of [26], Transductive Transfer Regularized Least-Squares Regression (TTRLSR) was proposed to solve the cross-database facial expression recognition problem. The core idea is to learn a discriminant subspace and use the source domain with complete data information and the dataset selected from the target domain sample to predict the facial expression category in the target domain. Chu et al. [27, 28] discovered a new method, selective transfer machine (STM), which can use target samples to learn a set of weights of the source samples, so that the weighted source samples and the target samples have the same or similar characteristic distribution. Then the learned classifier is used to predict the information of the target domain samples. In addition, Yan et al. [10] innovatively came up with Unsupervised Domain-Adaptive Dictionary Learning (UDADL), with the aim of learning a dictionary from the samples of source domain and target domain and solving the problem of cross-database facial expression recognition by using the basic idea of common subspace. In the Domain-Adaptive Subspace Learning (DoSL) model [29], the source

and target domain samples are mapped into the common subspace by learning a mapping matrix, and the regenerated source and target domain samples follow the same or similar feature distributions.

In the early research of cross-database speech emotion recognition, a series of good methods appeared. For example, Hassan et al. [30] proposed to use the Importance weighted support vector machine (IW-SVM) method to deal with the challenge of cross-database speech emotion recognition. The method was developed by combining three domain adaptive methods; namely, kernel mean matching (KMM) [31], Kullback–Leibler importance estimation procedure (KLIEP) [32], and unconstrained least-squares importance fitting (uLSIF) [33] were combined. It is used to alleviate the problem that the sample feature distribution of source domain and target domain does not match. Zong et al. [34] introduced the domain-adaptive least-squares regression (DALSR) method to learn a regression coefficient matrix using the source domain sample information and the auxiliary set selected from the target database. At the same time, the mean and variance of the speech sample in the source and target domain after regression are also constrained to make the sample in the two domains have similar feature distributions.

3. Subspace Learning and Joint Distribution Adaptation

3.1. Problem Definition. Assume that n and m denote the number of samples in the source and target domains, respectively, d represents the dimensionality of the micro-expression feature vector, and $X_s \in \mathbb{R}^{d \times n}$ and $X_t \in \mathbb{R}^{d \times m}$ are the microexpression samples in the source and target domains. Given a labeled source domain $\Omega_s = \{(x_j, y_j)\}_{j=1}^n$ and an unlabeled target domain $\Omega_t = \{x_i\}_{i=1}^m$, assume that the feature space and the category space are the same; that is, $\mathcal{X}_s = \mathcal{X}_t$ and $\mathcal{Y}_s = \mathcal{Y}_t$. The margin distribution is different $P_s(x_s) \neq P_t(x_t)$. Moreover, the mapping relationship is not the same in the source and target domains.

For better cross-database MER, we fuse data distribution and subspace learning methods to reduce domain differences. By mapping the source and target domains to the corresponding subspaces, we then maximize the target domain sample variance and preserve the source domain discriminative features while minimizing the two subspace differences and the subspace data distribution differences. The details are described next.

3.2. Subspace Learning and Joint Distribution Adaptation. Since the source domain contains not only sample information but also label information, we should make the best use of the discriminatory features of the source domain to achieve the best possible differentiation between different classes. By using the minimization of intraclass scatter and the maximization of interclass scatter, similar samples in the source domain can be close to each other, while different samples can be far away from each other as far as possible,

thus facilitating a learning algorithm with good performance.

$$\min_Y \text{Tr}(Y^T S_w Y), \quad (1)$$

$$S_w = \sum_{c=1}^C X_s^{(c)} H_s^{(c)} (X_s^{(c)})^T, \quad (2)$$

where S_w is the within-class scatter matrix, $H_s^{(c)} = I_s^{(c)} - 1/n_c \mathbf{1}_s^{(c)} (\mathbf{1}_s^{(c)})^T$ is the center matrix of class c in the source domain, $X_s^{(c)} \in \mathbb{R}^{d \times n_c}$, n_c is the number of class c , $\mathbf{1}_s \in \mathbb{R}^{n_c}$ is the row vector of unit 1, and $I_s^{(c)} \in \mathbb{R}^{n_c \times n_c}$ is the identity matrix.

$$\max_Y \text{Tr}(Y^T S_b Y), \quad (3)$$

$$S_b = \sum_{c=1}^C n_c (\mu_j - \mu) (\mu_j - \mu)^T, \quad (4)$$

where S_b is the interclass scatter matrix, μ_j is the sample mean of class c in the source domain, and μ is the sample mean of the source domain.

In order to preserve the data attributes to the classification of the target domain as much as possible, it is necessary to project the feature dimensions onto the relevant dimensions and therefore maximizing the target domain variance in the relevant subspace [35].

Maximize the target variance as follows:

$$\text{Tr}(Z^T X_t H_t X_t^T Z), \quad (5)$$

where $H_t = I_t - 1/m \mathbf{1}_t \mathbf{1}_t^T$ is the centering matrix and $\mathbf{1}_t \in \mathbb{R}^m$ is a row vector of unit 1.

Similar to the principle of data-centric domain adaptive methods, we compare the data differences between two domains by MMD distance to minimize the maximum mean difference between the source and target domains in k -dimensional embedding and further improve the marginal distribution differences.

$$\left\| \frac{1}{n} \sum_{x_j \in X_s} Y^T x_j - \frac{1}{m} \sum_{x_i \in X_t} Z^T x_i \right\|_F^2. \quad (6)$$

Because the label of the target domain is unknown, the conditional distribution of the target domain cannot be obtained directly, so the method proposed by [16] is usually used. Pseudo labels are obtained by using the classifier in the source domain to predict directly on the target domain. Since some pseudo labels may be incorrect and thus lead to poor recognition, an iterative approach is used to continuously improve the classification accuracy and thus reduce the difference between the two conditional distributions. The conditional distribution difference is

$$\sum_{c=1}^C \left\| \frac{1}{n_c} \sum_{x_j \in X_s^{(c)}} Y^T x_j - \frac{1}{m_c} \sum_{x_i \in X_t^{(c)}} Z^T x_i \right\|_F^2. \quad (7)$$

Therefore, by integrating equations (6) and (7), the distribution difference between source domain and target domain can be obtained:

$$\min_{Y,Z} \text{Tr} \left(\begin{bmatrix} Y^T & Z^T \end{bmatrix} \begin{bmatrix} X_s \left(M_0 + \sum_{c=1}^C M_c \right) X_s^T & X_s \left(M_0 + \sum_{c=1}^C M_c \right) X_t^T \\ X_t \left(M_0 + \sum_{c=1}^C M_c \right) X_s^T & X_t \left(M_0 + \sum_{c=1}^C M_c \right) X_t^T \end{bmatrix} \begin{bmatrix} Y \\ Z \end{bmatrix} \right). \quad (8)$$

To simplify equation (8), we write it as follows:

$$\min_{Y,Z} \text{Tr} \left[\begin{bmatrix} Y^T & Z^T \end{bmatrix} \begin{bmatrix} W_{s,s} & W_{s,t} \\ W_{t,s} & W_{t,t} \end{bmatrix} \begin{bmatrix} Y \\ Z \end{bmatrix} \right], \quad (9)$$

where

$$W_{s,s} = X_s \left(M_0 + \sum_{c=1}^C M_c \right) X_s^T,$$

$$(M_c)_{ji} = \begin{cases} \frac{1}{n_c^2}, & x_i, x_j \in X_s^{(c)}, \\ \frac{1}{m_c^2}, & x_i, x_j \in X_t^{(c)}, \\ \frac{1}{m_c n_c}, & \begin{cases} x_j \in X_s^{(c)}, x_i \in X_t^{(c)} \\ x_j \in X_t^{(c)}, x_i \in X_s^{(c)} \end{cases}, \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

$$(M_0)_{ji} = \begin{cases} \frac{1}{n^2} \mathbf{1}_s \mathbf{1}_s^T, & x_i, x_j \in X_s, \\ \frac{1}{m^2} \mathbf{1}_t \mathbf{1}_t^T, & x_i, x_j \in X_t, \\ -\frac{1}{mn} \mathbf{1}_s \mathbf{1}_t^T, & x_j \in X_s, x_i \in X_t, \\ -\frac{1}{mn} \mathbf{1}_t \mathbf{1}_s^T, & x_i \in X_t, x_j \in X_s. \end{cases}$$

Unlike the subspace learning method SA, which requires finding an alignment matrix and aligning the source and target domain subspaces to reduce the distance between the source and target domains, this method is mapped to their respective subspaces, and we optimize both Y and Z while making the two subspaces closer together.

$$\min_{Y,Z} \|Y - Z\|_F^2. \quad (11)$$

3.3. Optimization Problem. By integrating the five equations (1), (3), (5), (9), and (11), we can get the final SLJDA method:

$$\max \frac{\gamma(5) + \beta(3)}{\lambda(11) + (9) + \beta(1)}, \quad (12)$$

where λ , β , and γ are the hyperparameters by $\text{Tr}(Z^T Z)$ to constrain Z . Furthermore, the two coupled projections Y and Z are obtained using the following optimization function calculated:

$$\max_{Y,Z} \frac{\text{Tr} \left(\begin{bmatrix} Y^T & Z^T \end{bmatrix} \begin{bmatrix} \beta S_b & 0 \\ 0 & \gamma S_t \end{bmatrix} \begin{bmatrix} Y \\ Z \end{bmatrix} \right)}{\text{Tr} \left(\begin{bmatrix} Y^T & Z^T \end{bmatrix} \begin{bmatrix} W_{s,s} + \lambda I + \beta S_w & W_{s,t} - \lambda I \\ W_{t,s} - \lambda I & W_{t,t} + (\lambda + \gamma) I \end{bmatrix} \begin{bmatrix} Y \\ Z \end{bmatrix} \right)}. \quad (13)$$

Since $\begin{bmatrix} Y \\ Z \end{bmatrix}$ can be stretched without changing the final result, we can simply keep the bottom constant and only calculate the top. So the final optimization problem becomes

$$\max_{Y,Z} \text{Tr} \left(\begin{bmatrix} Y^T & Z^T \end{bmatrix} \begin{bmatrix} \beta S_b & 0 \\ 0 & \gamma S_t \end{bmatrix} \begin{bmatrix} Y \\ Z \end{bmatrix} \right) \\ \text{s.t. } \text{Tr} \left(\begin{bmatrix} Y^T & Z^T \end{bmatrix} \begin{bmatrix} W_{s,s} + \lambda I + \beta S_w & W_{s,t} - \lambda I \\ W_{t,s} - \lambda I & W_{t,t} + (\lambda + \gamma) I \end{bmatrix} \begin{bmatrix} Y \\ Z \end{bmatrix} \right) = 1. \quad (14)$$

Denote $\Phi = \text{diag}(\lambda_1, \dots, \lambda_d)$ as the Lagrange multiplier, and then Lagrange function for equation (14) is

$$L = \text{Tr} \left(\begin{bmatrix} Y^T & Z^T \end{bmatrix} \begin{bmatrix} \beta S_b & 0 \\ 0 & \gamma S_t \end{bmatrix} \begin{bmatrix} Y \\ Z \end{bmatrix} \right) \\ + \text{Tr} \left(\left(\begin{bmatrix} Y^T & Z^T \end{bmatrix} \begin{bmatrix} W_{s,s} + \lambda I + \beta S_w & W_{s,t} - \lambda I \\ W_{t,s} - \lambda I & W_{t,t} + (\lambda + \gamma) I \end{bmatrix} \begin{bmatrix} Y \\ Z \end{bmatrix} - I \right) \Phi \right). \quad (15)$$

Setting the derivative $\partial L / \partial A = 0$, we can obtain

$$\begin{bmatrix} \beta S_b & 0 \\ 0 & \gamma S_t \end{bmatrix} A = \begin{bmatrix} W_{s,s} + \lambda I + \beta S_w & W_{s,t} - \lambda I \\ W_{t,s} - \lambda I & W_{t,t} + (\lambda + \gamma) I \end{bmatrix} A \Phi, \quad (16)$$

where $\begin{bmatrix} Y \\ Z \end{bmatrix} = A$, and the matrix composed of eigenvectors corresponding to d minimum eigenvalues obtained by the

generalized value decomposition of equation (16) is A . Then the subspaces Y and Z are obtained.

4. Experiments

4.1. Microexpression Database. In this chapter, we conduct wide-ranging experiments in two popularly known databases, CASMEII and SMIC, to evaluate the performance of this method in a comprehensive manner. One of them, CASMEII [36], established by the Institute of Psychology of the Chinese Academy of Sciences, requires subjects not only to watch videos with large mood swings but more importantly to simultaneously attempt to mask their emotions. At the same time, the subjects' expressions were recorded at a frame rate of 200 fps without watching the video. With this elicitation mechanism, 247 video sequences from 26 individuals were acquired. Seven microexpression categories (*Happy, Surprised, Disgusted, Depressed, Sad, Scared, Other*) were included within this dataset. The SMIC dataset [37, 38] was established by the University of Oulu, Finland, using a similar induction mechanism to ensure the reliability of the data. The database contains 164 video sequences of 16 individuals (10 males and 6 females) containing 3 categories of microexpressions categories (*Positive, Surprised, Negative*). SMIC contains three datasets SMIC (HS, VIS, and NIR). The SMIC (VIS) and SMIC (NIR) were recorded in the Visual Identity System (VIS) and Near Infrared (NIR) environments, respectively, using a camera with a frame rate of 25 FPS. The results were also recorded by HS using a 100 FPS high-speed camera, and the three will be used as separate datasets in this experiment. As shown above, the SMIC and CASMEII databases contain inconsistent microexpression categories. To face this problem, we selected and reannotated the categories in CASMEII. Specifically, firstly, the *Happy* sample is changed to *Positive*, the *Surprised* sample is kept unchanged, then the *Other* category is deleted, and finally the *Disgusted* and *Depressed* samples correspond to *Negative*. The details are shown in Table 1.

4.2. Experimental Setup. Regarding SMIC (VIS, HS, and NIR) and CASMEII, we designed 24 sets of cross-database microexpression experiments. Among them, no. 1–no. 6 and no. 13–no. 18 are treated as the first type of experiment, which is labeled as type-1, and no. 7–no. 12 and no. 19–no. 24 are treated as the second type of experiment, which is labeled as type-2. We use LBP-TOP [39] and HIGO-TOP [40] spatiotemporal descriptors for feature extraction. For the former, the neighborhood radius R and neighborhood point P are set to 1 and 8, respectively. HIGO-TOP has only one important parameter P , which is set to 8. Whether it is LBP-TOP or HIGO-TOP, the feature vector subsets of each plane are sequentially concatenated to form a supervector, that is, to form a microexpression facial feature vector.

As we can see from Table 1, the CASMEII and SMIC databases have extremely unbalanced samples for each category. Therefore, in addition to the weighted average recall (WAR), we also used unweighted average recall (UAR) [41] as the measurement criterion to compare the

TABLE 1: Sample composition of the reconstituted CASMEII and SMIC databases.

Category	Database			
	CASMEII	SMIC (HS)	SMIC (NIR)	SMIC (VIS)
Positive	32	51	23	23
Surprise	25	43	20	20
Negative	91	70	28	28
Total	148	164	71	71

performance of various methods in the CDMER experiment. WAR is the rate of correctness. UAR is the average of the recognition rates calculated for each category of samples, regardless of the number of samples in each category. In summary, the combined WAR and UAR provide a fairer measure of the true performance of a method when the categories are unbalanced.

In order to show the performance of SLJDA, we choose other domain adaptation methods which have better performance in cross-database recognition to carry out comparative experiments. These methods include Support Vector Machine (SVM) [42], Domain Regeneration in the original Feature Space with unchanged Target domain (DRFS-T), Domain Regeneration in the original Label Space (DRLS), Target Sample Re-Generator (TSRG), SA, Joint Distribution Analysis (JDA), and Transfer Joint Matching (TJM). Parameter settings of all the above methods in the CDMER experiment are shown as follows:

- (1) For SVM, set $C = 1$ (linear), and use it as a traditional method. The results of experiments conducted directly with SVM will be used as a baseline method for comparison.
- (2) DRFS-T, DRLS, and TSRG involve two important parameters, that is, λ and μ . Following this work, we search for the best values of these two parameters in $\lambda \in [0.001, 0.01, 0.1, 1, 10, 100, 1000]$ and $\mu \in [0.001: 0.001: 0.009, 0.01: 0.01: 0.09, 0.1: 0.1: 1, 2: 10]$.
- (3) We set $\lambda \in [0.1, 1, 10]$ according to the author's suggestion in [18], and JDA follows the original design of $\lambda \in [0.01, 0.1, 1, 10, 100]$. In addition, optimal dimensionality reduction dimension $d \in [10: 10: \max]$. For SA, we will traverse all possible dimensions d , that is, search $[1: 1: \max]$.
- (4) Finally, with respect to SLJDA in this paper, we choose β from $[0.00005: 0.00001: 0.00009, 0.0001: 0.0001: 0.0009, 0.001: 0.001: 0.009, 0.01: 0.01: 0.09, 0.1: 0.1: 1, 2: 10]$, for $d \in [20: 10: \max]$.

4.3. Analysis of Experimental Results. Tables 2–5 show the results compared to the current better performing cross-database methods, and the bolded ones are the best performing methods in each group of experiments. Observing the table, we can find that the SLJDA method performs best in 8 out of 12 groups of experiments. In addition, we re-performed feature extraction with HIGO and compared it

TABLE 2: Type-1 experimental results of unsupervised CDMER based on LBP-TOP feature extraction (WAR).

No.	Task	SVM	DRLS	DRFS-T	TSRG	SLJDA
1	CASMEII-SMIC (HS)	44.51	54.27	43.90	45.73	57.93
2	CASMEII-SMIC (NIR)	62.16	79.73	70.95	68.92	54.73
3	CASMEII-SMIC (VIS)	35.21	54.93	32.39	40.85	56.34
4	SMIC (HS)-CASMEII	33.78	63.51	54.73	62.16	58.11
5	SMIC (NIR)-CASMEII	38.03	56.34	35.21	45.07	73.24
6	SMIC (VIS)-CASMEII	35.14	70.95	52.03	61.49	57.43

TABLE 3: Type-1 experimental results of unsupervised CDMER based on LBP-TOP feature extraction (UAR).

No.	Task	SVM	DRLS	DRFS-T	TSRG	SLJDA
1	CASMEII-SMIC (HS)	35.96	53.71	34.88	37.81	55.46
2	CASMEII-SMIC (NIR)	34.67	73.46	51.12	47.71	46.77
3	CASMEII-SMIC (VIS)	36.67	55.15	33.33	43.33	56.30
4	SMIC (HS)-CASMEII	40.32	66.22	55.86	35.05	54.11
5	SMIC (NIR)-CASMEII	40.00	36.67	58.94	48.33	72.27
6	SMIC (VIS)-CASMEII	49.94	54.70	49.16	33.33	67.85

TABLE 4: Type-2 experimental results of unsupervised CDMER based on LBP-TOP feature extraction (WAR).

No.	Task	SVM	DRLS	DRFS-T	TSRG	SLJDA
7	SMIC (HS)-SMIC (NIR)	61.97	73.24	78.87	73.24	64.79
8	SMIC (NIR)-SMIC (HS)	46.34	60.98	57.32	58.54	60.98
9	SMIC (HS)-SMIC (VIS)	76.06	88.73	78.87	81.69	95.77
10	SMIC (VIS)-SMIC (HS)	54.88	64.63	59.15	56.71	65.85
11	SMIC (NIR)-SMIC (VIS)	78.87	81.69	80.28	77.46	85.92
12	SMIC (VIS)-SMIC (NIR)	66.20	77.46	78.87	77.46	81.69

TABLE 5: Type-2 experimental results of unsupervised CDMER based on LBP-TOP feature extraction (UAR).

No.	Task	SVM	DRLS	DRFS-T	TSRG	SLJDA
7	SMIC (HS)-SMIC (NIR)	60.52	72.32	78.34	72.57	65.13
8	SMIC (NIR)-SMIC (HS)	58.02	62.85	58.02	59.87	62.97
9	SMIC (HS)-SMIC (VIS)	77.17	88.31	79.50	83.01	95.21
10	SMIC (VIS)-SMIC (HS)	59.15	67.74	60.47	60.41	66.73
11	SMIC (NIR)-SMIC (VIS)	77.86	82.23	79.53	77.15	86.36
12	SMIC (VIS)-SMIC (NIR)	65.24	76.06	78.29	75.63	82.02

with the currently popular domain adaptive methods. As shown in Tables 6 and 7, more than half of the experimental results show that this method has a clear advantage over other methods.

Observing the table, we can find that the experimental results of no. 7–no. 12 (type-2) and no. 19–no. 24 (type-2) are better than those of no. 1–no. 6 (type-1) and no. 13–no. 18 (type-1). The analysis may be due to the fact that the differences between CASMEII and SMIC samples are greater than differences between three datasets of SMIC. This is because the SMIC (HS), SMIC (VIS), and SMIC (NIR) only differ in the camera pixels taken between them. However, between CASMEII and SMIC, the filming equipment, recording environment, ethnicity of the subjects, and so on were different. It is worth noting that SMIC (VIS) experimental results are better than SMIC (NIR) experimental results under the same conditions, both as source and target domain datasets, for example, no. 3 and no. 5. Combining 4.1, we analyze that this may be due to the image quality difference problem.

TABLE 6: Type-1 experimental results of unsupervised CDMER based on HIGO-TOP feature extraction (WAR).

No.	Task	SA	JDA	TJM	SLJDA
13	CASMEII-SMIC (HS)	61.95	53.05	56.1	53.05
14	CASMEII-SMIC (NIR)	66.89	43.92	45.95	56.76
15	CASMEII-SMIC (VIS)	50.7	57.75	52.11	53.52
16	SMIC (HS)-CASMEII	48.65	45.95	45.27	54.73
17	SMIC (NIR)-CASMEII	57.75	59.15	61.97	66.20
18	SMIC (VIS)-CASMEII	49.32	46.62	47.3	58.78

TABLE 7: Type-2 experimental results of unsupervised CDMER based on HIGO-TOP feature extraction (WAR).

No.	Task	SA	JDA	TJM	SLJDA
19	SMIC (HS)-SMIC (NIR)	73.24	71.83	73.24	70.42
20	SMIC (NIR)-SMIC (HS)	59.15	60.98	49.39	61.59
21	SMIC (HS)-SMIC (VIS)	77.46	92.96	94.37	94.37
22	SMIC (VIS)-SMIC (HS)	58.54	60.37	59.76	65.24
23	SMIC (NIR)-SMIC (VIS)	84.51	74.65	76.06	87.32
24	SMIC (VIS)-SMIC (NIR)	74.65	78.87	77.46	77.46

In addition, an interesting phenomenon prevails in all the methods covered in this paper; when CASMEII and HS are used as target domains, the difference between WAR and UAR of the former is greater than that of the latter. For example, in Tables 2–5, the difference between WAR and UAR of no. 5 and no. 6 is significantly larger than the difference between no. 8 and no. 10. Looking at Table 1, it can be observed that the CASMEII dataset and the SMIC (HS) dataset are unbalanced in terms of microexpression categories, with the former having a majority of negative emotions (91/148). At the same time, the proportion of negative emotions in SMIC (HS) (70/164) is smaller than that of CASMEII. Analysis of this reason may be due to the imbalance of categories in the dataset. Further observation of both SMIC (NIR) and SMIC (VIS) databases showed consistent proportions of microexpression categories. When SMIC (NIR) or SMIC (VIS) is the target domains (no. 7 or no. 11), the difference between WAR and UAR is smaller than that between SMIC (HS) and CASMEII, which are also the target domains. This further proves that the category imbalance of microexpression database has a certain influence on MER.

5. Conclusion and Discussion

In this paper, we integrate data distribution methods with subspace learning methods from practical situations as a way to reduce the differences between source and target domains. That is, the SLJDA method is used to perform unsupervised cross-database microexpression recognition. Extensive experiments are conducted on CASMEII and SMIC databases, and the experimental results show that this method significantly outperforms other state-of-the-art domain adaptive methods for unsupervised CDMER tasks. However, there are still some issues to be investigated. As shown above, category imbalance in microexpression database is an important factor affecting the results of CDMER. Therefore, to improve the recognition results, we need to work on reducing the impact caused by the database category imbalance.

Data Availability

The data used to support this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Y. Zhang and Y. Liu jointly designed the study. Y. Liu collected and analyzed the data. G. Li and H. Peng reviewed and edited the manuscript.

Acknowledgments

This work was supported by the Key Research Project of Henan Higher Education Institution (Project no. 17B440001), Henan University of Science and Technology

Enterprise Commissioning Project (JG017), and Henan Provincial Science and Technology Tackling Program Grant Project 212102210504.

References

- [1] P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*, WW Norton & Company, New York, NY, USA, 2009.
- [2] P. Ekman, "Darwin, deception, and facial expression," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, pp. 205–221, 2003.
- [3] J. M. Gottman and R. W. Levenson, "A two-factor model for predicting when a couple will divorce: exploratory analyses using 14-year longitudinal data," *Family Process*, vol. 41, no. 1, pp. 83–96, 2002.
- [4] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: automatic recognition of student engagement from facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 2014.
- [5] L. Dacre Pool and P. Qualter, "Improving emotional intelligence and emotional self-efficacy through a teaching intervention for university students," *Learning and Individual Differences*, vol. 22, no. 3, pp. 306–312, 2012.
- [6] X. Li, J. Li, Y. Liu, Z. Ding, and A. Nallanathan, "Residual transceiver hardware impairments on cooperative NOMA networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 680–695, 2020.
- [7] X. Li, M. Zhao, M. Zeng et al., "Hardware impaired ambient backscatter NOMA system: reliability and security," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2723–2736, 2021.
- [8] X. Li, M. Zhao, Y. Liu, L. Li, Z. Ding, and A. Nallanathan, "Secrecy analysis of ambient backscatter NOMA systems under I/Q imbalance," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12286–12290, 2020.
- [9] X. Li, Q. Wang, M. Liu et al., "Cooperative wireless-powered NOMA relaying for B5G IoT networks with hardware impairments and channel estimation errors," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5453–5467, 2020.
- [10] K. Yan, W. Zheng, Z. Cui, and Y. Zong, "Cross-database facial expression recognition via unsupervised domain adaptive dictionary learning," *International Conference on Neural Information Processing*, Springer, vol. 9948, pp. 427–434, Berlin, Germany, 2016.
- [11] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [12] Y. Zong, W. Zheng, X. Huang, J. Shi, Z. Cui, and G. Zhao, "Domain regeneration for cross-database micro-expression recognition," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2484–2498, 2018.
- [13] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.
- [14] J. Wang, Y. Chen, L. Hu, X. Peng, and P. S. Yu, "Stratified transfer learning for cross-domain activity recognition," in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, pp. 1–10, Athens, Greece, March 2018.
- [15] K. Saito, Y. Ushiku, and T. Harada, "Asymmetric tri-training for unsupervised domain adaptation," in *Proceedings of the 34th International Conference on Machine Learning*, pp. 2988–2997, Sydney, Australia, February 2017.

- [16] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2200–2207, Sydney, Australia, December 2013.
- [17] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *Proceedings of the IEEE International Conference on Data Mining*, pp. 1129–1134, New Orleans, LA, USA, November 2017.
- [18] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1410–1417, Columbus, OH, USA, June 2014.
- [19] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 120–128, Sydney, Australia, July 2006.
- [20] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2960–2967, Sydney, NSW, Australia, December 2013.
- [21] B. Sun and K. Saenko, "Deep coral: correlation alignment for deep domain adaptation," *European Conference on Computer Vision*, Springer, Berlin, Germany, pp. 443–450, 2016.
- [22] B. Sun and K. Saenko, "Subspace distribution alignment for unsupervised domain adaptation," in *Proceedings of the British Machine Vision Conference 2015*, vol. 4, pp. 24–31, Lowell, MA, USA, September 2015.
- [23] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2066–2073, Providence, RI, USA, June 2012.
- [24] L. Li, X. Zhou, Y. Zong et al., "Unsupervised cross-database micro-expression recognition using target-adapted least-squares regression," *IEICE Transactions on Information and Systems*, vol. E102.D, no. 7, pp. 1417–1421, 2019.
- [25] R. Zhu, T. Zhang, Q. Zhao, and Z. Wu, "A transfer learning approach to cross-database facial expression recognition," in *Proceedings of the International Conference on Biometrics*, pp. 293–298, Phuket, Thailand, May 2015.
- [26] W. Zheng and X. Zhou, "Cross-domain color facial expression recognition using transductive transfer subspace learning," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 21–37, 2016.
- [27] W. S. Chu, F. De La Torre, and J. F. Cohn, "Selective transfer machine for personalized facial action unit detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3515–3522, Portland, OR, USA, June 2013.
- [28] W.-S. Chu, F. De La Torre, and J. F. Cohn, "Selective transfer machine for personalized facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 529–545, 2017.
- [29] N. Liu, Y. Zong, B. Zhang et al., "Unsupervised cross-corpus speech emotion recognition using domain-adaptive subspace learning," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5144–5148, Calgary, Canada, April 2018.
- [30] A. Hassan, R. Damper, and M. Niranjana, "On acoustic emotion recognition: compensating for covariate shift," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1458–1468, 2013.
- [31] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Scholkopf, "Correcting sample selection bias by unlabeled data," in *Advances in Neural Information Processing Systems*, pp. 601–608, MIT Press, Cambridge, MA, USA, 2007.
- [32] M. Sugiyama, T. Suzuki, S. Nakajima et al., "Direct importance estimation with model selection and its application to covariate shift adaptation," *Advances in Neural Information Processing Systems*, vol. 20, pp. 1433–1440, 2007.
- [33] T. Kanamori, S. Hido, and M. Sugiyama, "A least-squares approach to direct importance estimation," *Journal of Machine Learning Research*, vol. 10, no. 48, pp. 1391–1445, 2009.
- [34] Y. Zong, W. Zheng, T. Zhang, and X. Huang, "Cross-corpus speech emotion recognition based on domain-adaptive least-squares regression," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 585–589, 2016.
- [35] J. Zhang, W. Li, and P. Ogunbona, "Joint geometrical and statistical alignment for visual domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5150–5158, Honolulu, HI, USA, May 2017.
- [36] W. J. Yan, Q. Wu, Y. J. Liu, S.-J. Wang, and X. Fu, "CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces," in *Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–7, Shanghai, China, April 2013.
- [37] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: inducement, collection and baseline," in *Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–6, Shanghai, China, April 2013.
- [38] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *Proceedings of the International Conference on Computer Vision*, pp. 1449–1456, Barcelona, Spain, November 2011.
- [39] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.
- [40] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 886–893, San Diego, CA, USA, June 2005.
- [41] B. Schuller, B. Vlasenko, F. Eyben et al., "Cross-corpus acoustic emotion recognition: variances and strategies," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 119–131, 2011.
- [42] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.