# Toward Bridging Microexpressions From Different Domains

Yuan Zong, *Member, IEEE*, Wenming Zheng, *Senior Member, IEEE*, Zhen Cui, *Member, IEEE*, Guoying Zhao, *Senior Member, IEEE*, and Bin Hu, *Senior Member, IEEE*

***Abstract*—Recently, microexpression recognition has attracted a lot of researchers' attention due to its challenges and valuable applications. However, it is noticed that currently most of the existing proposed methods are often evaluated and tested on the single database and, hence, this brings us a question whether these methods are still effective if the training and testing samples belong to different domains, for example, different microexpression databases. In this case, a large feature distribution difference may exist between training (source) and testing (target) samples and, hence, microexpression recognition tasks would become more difficult. To solve this challenging problem, that is, cross-domain microexpression recognition, in this paper, we propose an effective method consisting of an auxiliary set selection model (ASSM) and a transductive transfer regression model (TTRM). In our method, an ASSM is designed to automatically select an optimal set of samples from the target domain to serve as the auxiliary set, which is used for subsequent TTRM training. As for TTRM, it aims at bridging the feature distribution gap between the source and target domains by learning a joint regression model with the source domain samples and the auxiliary set selected from the target domain. We evaluate the proposed TTRM plus ASSM by extensive cross-domain microexpression recognition experiments on SMIC and CASME II databases. Compared with the recent state-of-the-art domain adaptation methods, our proposed method has a more satisfactory performance in dealing with the cross-domain microexpression recognition tasks.**

***Index Terms*—Cross-domain microexpression recognition, domain adaptation (DA), microexpression recognition, transductive transfer regression, transfer learning.**

Manuscript received March 19, 2018; revised August 1, 2018 and February 8, 2019; accepted April 16, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1305200, in part by the National Basic Research Program of China under Grant 2015CB351704, in part by the National Natural Science Foundation of China under Grant 61572009, Grant 61632014, Grant 61802058, and Grant 6181101568, in part by the Fundamental Research Funds for the Central Universities under Grant 2242018K3DN01 and Grant 2242019K40047, in part by the China Scholarship Council, in part by the Tencent AI Lab Rhino-Bird Focused Research Program under Grant JR201922, in part by the Academy of Finland, in part by the Tekes Fidipro Program, and in part by Infotech Oulu. This paper was recommended by Associate Editor M. Zhang. *(Corresponding authors: Wenming Zheng; Bin Hu.)*

Y. Zong is with the Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China, and also with the Center for Machine Vision and Signal Analysis, Faulty of Information Technology and Electrical Engineering, University of Oulu, 90014 Oulu, Finland (e-mail: xhzongyuan@seu.edu.cn).

W. Zheng is with the Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China (e-mail: wenming_zheng@seu.edu.cn).

Z. Cui is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: zhen.cui@njust.edu.cn).

G. Zhao is with the Center for Machine Vision and Signal Analysis, Faulty of Information Technology and Electrical Engineering, University of Oulu, 90014 Oulu, Finland (e-mail: guoying.zhao@oulu.fi).

B. Hu is with the Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China (e-mail: bh@lzu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2019.2914512

## I. Introduction

MICROEXPRESSION recognition aims at accurately detecting the human beings' inner true emotional states which they try to conceal from the facial video clips [1]. It has been one of the most attractive research issues among affective computing, human behavior analysis, and pattern-recognition fields since it has many valuable practical applications, such as clinical diagnosis [2], interrogation [3], and security [4]. Compared with the ordinary dynamic facial expressions [5], microexpressions have two important characteristics, that is, lower intensity and shorter duration, which makes microexpression recognition become a very difficult and challenging task. Nevertheless, in recent years, many researchers have been devoted to investigate microexpression recognition and proposed lots of effective methods. In the work of [1], Pfister *et al.* first proposed using a local binary pattern from three orthogonal planes (LBP-TOPs) [6] to describe microexpressions and demonstrated its effectiveness in microexpression recognition tasks. Subsequently, various techniques are applied to enhance the performance of the LBP-TOP descriptor such that it can be more suitable for microexpression recognition. For example, reparameterization of the second-order Gaussian jet was used to improve LBP-TOP for describing microexpressions in the work of [7] and the proposed method achieved a more promising result than that of [1]. Wang *et al.* [8] proposed employing robust principal component analysis (RPCA) [9] to extract the background information from the image sequences of microexpression samples and then use the LBP-TOP extracted from the background information to describe the microexpressions. Recently, another novel descriptor called spatiotemporal local binary pattern with integration projection (STLBP-IP) was developed by Huang *et al.* [10] to describe microexpressions. More recently, some other new spatiotemporal descriptors, for example, TOP versions of completed local

quantized patterns (CLQP-TOP) [11], histogram of oriented gradients (HOGs-TOP) [12], and histogram of image gradient orientation (HIGO-TOP) [12], have been proposed for microexpression analysis. In addition to the LBP-TOP methodology, there are still other types of descriptors designed for microexpression recognition. The mean directional mean optical (MDMO) [13] and facial dynamics map (FDM) [14] are two representative-leading microexpression features among them.

Although the research of microexpression recognition has lasted for several years, it is noticed that nearly all of them focus on the scenario where the training and testing samples are from the same microexpression database. Usually, in this case, the training and testing samples can be thought to share the same feature distributions [15], [16]. In practice, however, such a condition may be broken because microexpression samples used for the training and testing stage would be quite different. For instance, they may be recorded by different equipment, under different environments, or the subjects of the training and testing microexpression samples belong to different ethnics. Consequently, it is necessary to consider whether the aforementioned microexpression recognition methods are still workable if the testing samples are different from the training samples. This thus leads to a more challenging microexpression recognition problem, namely, cross-domain microexpression recognition. For a convenient description in what follows, we will refer to the training samples as the source domain (samples) and the testing samples as the target domain (samples), respectively.

As a typical domain adaptation (DA) problem, the cross-domain microexpression recognition can be roughly divided into two cases according to the number of the provided labeled samples from the target domain, that is, the semisupervised case and unsupervised case [17]. Table I illustrates the detailed label information provided in these two cases. From this table, it is clear to see that the unsupervised case is more difficult and challenging than the semisupervised one because the label information of the target domain is entirely unknown. In this paper, we will focus on the unsupervised cross-domain microexpression recognition problem and propose a novel method whose basic idea is illustrated in Fig. 1. As Fig. 1 shows, the proposed method consists of two major steps, where each step corresponds to one model. In the first step, we develop an auxiliary set selection model (ASSM) to select an optimal subset from the unlabeled target domain to serve as the auxiliary set. In the second step, the selected auxiliary set is used together with the labeled source samples to jointly train a transductive transfer regression model (TTRM). By resorting to TTRM, we are able to narrow the feature distribution gap between the source and target domains existing in the original feature space.

It is worth mentioning that this paper is actually a follow-up to our previous work of [18], in which we proposed a transductive transfer subspace learning (TTSL) model to deal with the unsupervised cross-domain facial expression recognition problem. However, TTSL still has several limitations. First, TTSL and its relaxed version called transductive transfer regularized least squares regression (TTRLSR) proposed in [18]

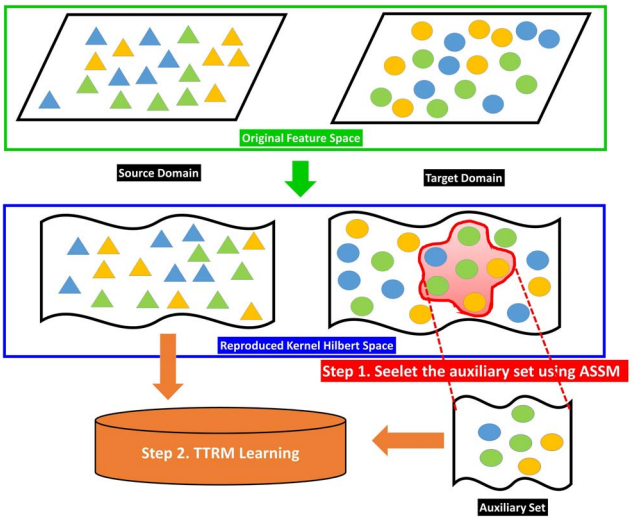


Fig. 1. Basic idea of the proposed method consisting of ASSM and TTRM for unsupervised cross-domain microexpression recognition.

TABLE I
CATEGORIZATION OF CROSS-DOMAIN MICRO-EXPRESSION RECOGNITION PROBLEM

| | Source Domain | Target Domain |
|---|---|---|
| Semi-supervised case | Labeled | Labeled (a few of samples) |
| Unsupervised case | Labeled | Unlabeled |

only consider the contributions of the facial local regions to distinguish different expressions with a binary manner, while their contributions are quite different in recognizing expressions. Second, TTSL needs to select an auxiliary set from the target domain in advance. So far, it is still an open question for TTSL to seek the elements and the optimal size of the auxiliary set from the target domain, which is very important to TTSL and needs to be solved urgently.

By taking the above limitations of TTSL into consideration, we generate the idea of ASSM and TTRM. Compared with TTRLSR, the proposed TTRM is not only one more relaxed and solvable version of TTSL but it also considers the specific contributions of different facial local regions to distinguish microexpressions. As for ASSM, it targets conquering the second point of the above limitations of TTSL work. Overall, different from our previous work of TTSL [18], this paper has the following three new contributions.

1) We propose a new relaxed and solvable version of TTSL called TTRM. In contrast to the TTRLSR method, TTRM is able to quantify the contributions of facial local regions in distinguishing different microexpressions whereas TTSL can only roughly determine whether each facial block is needed.
2) We propose a simple yet effective method called ASSM to solve the problem of choosing the optimal auxiliary set from the unlabeled data samples of the target domain, which is an important yet unsolved problem in the TTSL method.

3) We generalize the TTRM method from the feature space to a nonlinear reproduced kernel Hilbert space (RKHS) such that the TTRM model in RKHS can better fit the auxiliary data samples selected by ASSM and then we can also simplify the TTRM model.

The remainder of this paper is organized as follows. Section II briefly reviews the related works to unsupervised cross-domain microexpression recognition. Then, some preliminary works and backgrounds associated with the proposed method for unsupervised cross-domain microexpression recognition are introduced in Section III. Section IV describes our proposed TTRM plus ASSM (TTRM + ASSM)-based unsupervised cross-domain microexpression recognition method in detail. In Section V, extensive unsupervised cross-domain microexpression recognition experiments are conducted to evaluate our proposed method. Finally, this paper is concluded in Section VI.

## II. Related Works

In this section, we first briefly introduce the existing work of unsupervised cross-domain microexpression recognition. Then, we also review the recent progress of other modality (including speech emotion and facial expression)-based unsupervised cross-domain emotion recognition and DA in other applications, which is closely related to unsupervised cross-domain microexpression recognition.

### *A. Unsupervised Cross-Domain Micro-Expression Recognition*

Recently, the problem of cross-domain microexpression recognition has gradually drawn the researchers' attentions. In the work of [17], Zong *et al.* first investigated unsupervised cross-database microexpression recognition by proposing a target sample regenerator (TSRG) method. This method targets at learning a sample regenerator to regenerate the source and target microexpression samples such that the regenerated source and target samples would have the same or similar feature distributions. Then, we are able to train a classifier based on the labeled source microexpression samples and use it to predict the microexpression categories of the unlabeled regenerated target samples.

### *B. Other Modality-Based Unsupervised Cross-Domain Emotion Recognition*

As for unsupervised cross-domain speech emotion recognition, the first research may be traced to the work of [19], in which Schuller *et al.* investigated cross-corpus speech emotion recognition problem by using various normalization schemes to normalize the speech features. Since then, many researchers have focused on this interesting problem and proposed lots of effective methods. For instance, Deng *et al.* [20]–[22] designed a sequence of autoencoder-based unsupervised domain adaption methods, which learns a common representation of the speech features from source and target domains to bridge the two domains, for handling cross-corpus speech emotion recognition tasks. In the work of [23], Hassan *et al.* proposed an importance-weighted support vector machine (IW-SVM) to deal with cross-corpus speech emotion recognition problem. The importance weights of IW-SVM are learned by three typical DA methods, that is, kernel mean matching (KMM) [24], Kullback–Leibler importance estimation procedure (KLIEP) [25], and unconstrained least-squares importance fitting (uLSIF) [26].

For facial expression modality, lots of interesting works about unsupervised cross-domain facial expression recognition have also emerged in the recent years. For example, Chu *et al.* [27], [28] proposed a novel method called selective transfer machine (STM) to cope with the personalized facial action units detection problem. The key novelty of this paper is that STM can learn a set of weights for the corresponding source samples by utilizing the testing samples such that the classifier could also be suitable for target samples. In the work of [29], Sangineto *et al.* investigated the unsupervised cross-domain facial expression recognition problem and proposed a novel classifier parameter transfer model that directly transfers the knowledge about the parameters of source domain classifier to the classifier for target domain. Recently, Yan *et al.* [30] proposed an unsupervised domain-adaptive dictionary learning (UDADL) model to cope with the unsupervised cross-database facial expression recognition problem. UDADL aims at learning a common dictionary for source and target facial expressions such that the new representations with respect to the learned common dictionary share the same feature distributions.

### *C. Domain Adaptation in Other Applications*

Cross-domain emotion recognition, including microexpression recognition, is just a typical application field of DA methods. In fact, DA methods have great values in lots of applications in computer vision and multimedia analysis. For instance, the widely used baseline unsupervised DA method, geodesic flow kernel (GFK) [31], was originally proposed to solve cross-domain object image recognition. GFK aims to bridge two different domains and narrow their gaps with a well-designed GFK on a Grassmann manifold. Another example may be the action recognition field. For action recognition tasks, DA can provide a new and different solving angle. In the work of [32], Zhang *et al.* proposed a novel DA method called semisupervised image-to-video adaptation (IVA) to deal with the video action recognition problem and achieved satisfactory performance. IVA adapts the knowledge from images such that the action recognition in videos can be enhanced. DA can also be used to solve the image attribute prediction tasks, which is an interesting and meaningful vision problem. For example, Han *et al.* [33] proposed a DA framework called image attribute adaptation (IAA), which aims to automatically adapt the knowledge of attributes from well-defined auxiliary images to target images. Thus, IAA can assist in predicting appropriate attributes for target images.

## III. Preliminary

In this section, we introduce some preliminary knowledge that is needed and contribute to the derivation and

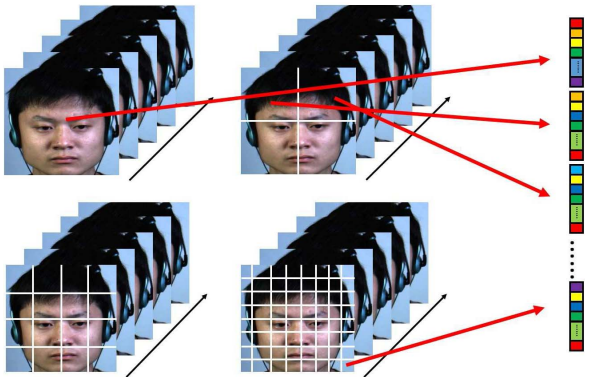

Fig. 2. Multiscale spatial division method for microexpression feature extraction.

understanding of the proposed TTRM + ASSM-based unsupervised cross-domain microexpression recognition method.

### A. Micro-Expression Feature Extraction

In the research of facial expression and microexpression analysis, it is an important step to extract the spatiotemporal descriptors from video clips to describe facial expression and microexpression samples. During this step, a spatial grid, for example, $8 \times 8$, is often used to divide the facial video clip into a few facial local regions in advance. Instead of directly using the grid with fixed size, Zhao and Pietikäinen [34] proposed employing a multiscale-division method consisting of multiple grids with different sizes to divide the facial expression video clips. Their work suggested that such combination of multiscale facial local regions provides more beneficial information to distinguish dynamic facial expressions. Similarly, this multiscale-division method also contributes to the descriptors, such as local binary pattern (LBP) [35], in describing the static image facial expression recognition [36]–[38]. Motivated by these works, in this paper, we employ this spatial-division scheme for microexpression feature extraction and choose four types of grids, that is, $1 \times 1$, $2 \times 2$, $4 \times 4$, and $8 \times 8$, for the scheme. An illustration of our feature extraction method used in this paper is shown in Fig. 2. Concretely, given a microexpression video clip $\mathcal{V}$, we first divide it into $M = 85$ facial local regions as shown in the example. Then, a specific spatiotemporal descriptor $\mathbf{x}_i (i = 1, \ldots, M)$, for example, LBP-TOP, is extracted from each facial local region, where $\mathbf{x}_i$ is a column vector. Finally, these spatiotemporal descriptors are concatenated, in turn, to compose a feature vector, which is denoted by $\mathbf{x}^{\mathcal{V}} = [\mathbf{x}_1^T, \ldots, \mathbf{x}_M^T]^T$, to describe this microexpression sample.

### B. Transductive Transfer Subspace Learning [18]

The TTSL model is originally proposed to deal with the unsupervised cross-domain facial expression recognition problem. In this paper, we will introduce TTSL by applying it to solve the unsupervised cross-domain microexpression recognition problem such that TTSL can be better understood from the viewpoint of our topic. Suppose we have $N_s$ source microexpression samples and $N_{t_{au}}$ auxiliary samples selected from the target microexpression database. Their corresponding feature matrices, which are extracted according to the proposed multiscale feature extraction scheme, are denoted by $\mathbf{X}^s = [\mathbf{X}_1^{s^T}, \ldots, \mathbf{X}_M^{s^T}]^T \in \mathbb{R}^{Md \times N_s}$ and $\mathbf{X}^{t_{au}} = [\mathbf{X}_1^{t_{au}^T}, \ldots, \mathbf{X}_M^{t_{au}^T}]^T \in \mathbb{R}^{Md \times N_{t_{au}}}$, where each column of the source and target feature matrices is the feature vector like $\mathbf{x}^{\mathcal{V}}$ shown in Section III-A, $M$ is the number of the divided facial local regions and $d$ is the dimension of the feature vector extracted from each facial local region, respectively. Let $\mathbf{Y}^s$ and $\mathbf{Y}^{t_{au}}$ be their corresponding label matrices and each column of them contains its corresponding microexpression sample's label information where all of its elements are binary and only the $c$th element is 1 if it belongs to the $c$th microexpression category. Then, we are able to formulate the TTSL model as the following optimization problem:

$$\min_{\mathbf{Y}^{t_{au}}, \mathbf{A}, \mathbf{B}_i, \omega_i, \mathbf{W}^s, \mathbf{W}^t} f_1^{\text{TTSL}} + \lambda_1 f_2^{\text{TTSL}} + \lambda_2 f_3^{\text{TTSL}}$$
$$\text{s.t. } \mathbf{l}_j^{t_{au}} \geq 0, \ \mathbf{1}^T \mathbf{y}_j^{t_{au}} = 1, \ \omega_i \in \{0, 1\} \quad (1)$$

where $\mathbf{y}_j^{t_{au}}$ is the $j$th column of the label matrix $\mathbf{Y}^{t_{au}}$ of the auxiliary samples. It is clear that the objective function of our TTSL model consists of three key terms, that is: 1) the loss function term $f_1^{\text{TTSL}}$; 2) the group sparse term $f_2^{\text{TTSL}}$; and 3) the difference elimination term $f_3^{\text{TTSL}}$. Next, we introduce them in detail.

*1) Loss Function Term $f_1^{TTSL}$:* The loss function term $f_1^{\text{TTSL}}$ aims at building the relationship between both source and auxiliary features and their corresponding microexpression label information and is defined as

$$f_1^{\text{TTSL}}(\mathbf{Y}^{t_{au}}, \mathbf{A}, \mathbf{B}_i, \omega_i) = \left\| [\mathbf{Y}^s, \mathbf{Y}^{t_{au}}] - \mathbf{A} \sum_{i=1}^{M} \omega_i \mathbf{B}_i^T [\mathbf{X}_i^s, \mathbf{X}_i^{t_{au}}] \right\|_F^2.$$

Note that $\mathbf{A}$ and $\mathbf{B}_i$ can be interpreted as the projection matrix which bridges the feature space and the label space and $\mathbf{B}_i$ is the $i$th matrix block of $\mathbf{B} = [\mathbf{B}_1^T, \ldots, \mathbf{B}_M^T]^T$. $\omega_i$ is a binary-weighted parameter and is designed to select the important facial local regions which have contributions to distinguish different microexpressions.

*2) Group Sparse Term $f_2^{TTSL}$:* As mentioned before, the target of introducing $\omega_i$ is to determine whether its corresponding facial block is needed for TTSL. To achieve this goal, the TTSL model adopts the L1-norm with respect to $\omega = [\omega_1, \ldots, \omega_M]^T$ to serve as the regularization term as follows:

$$f_2^{\text{TTSL}}(\omega_i) = \sum_{i=1}^{M} \omega_i.$$

*3) Difference Elimination Term $f_3^{TTSL}$:* $f_3^{\text{TTSL}}$ is a regularization term with respect to $\mathbf{B}$ and is designed to narrow the feature distribution gap between the source and target domains in the row space of $\mathbf{B}$. In the TTSL model, we employ the common subspace approach [39] as the $f_3^{\text{TTSL}}$ that has the following formulation:

$$f_3^{\text{TTSL}}(\mathbf{B}, \mathbf{W}^s, \mathbf{W}^t) = \left\| \mathbf{B}^T \mathbf{X}^s - \mathbf{B}^T \mathbf{X}^{t_{au}} \mathbf{W}^t \right\|_F^2 + \left\| \mathbf{B}^T \mathbf{X}^s \mathbf{W}^s - \mathbf{B}^T \mathbf{X}^{t_{au}} \right\|_F^2 + \lambda_3 \left( \left\| \mathbf{W}^s \right\|_1 + \left\| \mathbf{W}^t \right\|_1 \right)$$

where $\mathbf{W}^s$ and $\mathbf{W}^t$ are the linear combination coefficient matrices. The optimal projection matrix $\mathbf{B}$ should be the one which minimizes this term.

The optimization problem of the TTSL model does not have an effective solving method due to the strict binary constraint with respect to the weighted parameter $\omega_i$. Therefore, in the work of [18], we relax the original TTSL to a solvable version called the TTRLSR model by introducing a new variable $\tilde{\mathbf{B}}_i$ by satisfying $\tilde{\mathbf{B}}_i = \omega_i \mathbf{B}_i$. Thus, the optimization problem of TTSL in (1) would become solvable, which has the following formulation:

$$\begin{aligned} \min_{\mathbf{Y}^{t_{au}}, \mathbf{A}, \tilde{\mathbf{B}}_i, \mathbf{W}^s, \mathbf{W}^t} \quad & f_1^{\text{TTRLSR}} + \lambda_1 f_2^{\text{TTRLSR}} + \lambda_2 f_3^{TTRLSR} \\ \text{s.t.} \quad & \mathbf{y}_j^{t_{au}} \succeq 0, \ \mathbf{1}^T \mathbf{y}_j^{t_{au}} = 1 \end{aligned} \tag{2}$$

where

$$f_1^{\text{TTRLSR}}\left(\mathbf{Y}^{t_{au}}, \mathbf{A}, \tilde{\mathbf{B}}_i\right) = \left\| \left[\mathbf{Y}^s, \mathbf{Y}^{t_{au}}\right] - \mathbf{A} \sum_{i=1}^{M} \tilde{\mathbf{B}}_i^T \left[\mathbf{X}_i^s, \mathbf{X}_i^{t_{au}}\right] \right\|_F^2$$

$$f_2^{\text{TTRLSR}}\left(\tilde{\mathbf{B}}_i\right) = \sum_{i=1}^{M} \left\| \tilde{\mathbf{B}}_i \right\|_F$$

and

$$\begin{aligned} & f_3^{\text{TTRLSR}}\left(\tilde{\mathbf{B}}, \mathbf{W}^s, \mathbf{W}^t\right) \\ & = \left\| \tilde{\mathbf{B}}^T \mathbf{X}^s - \tilde{\mathbf{B}}^T \mathbf{X}^{t_{au}} \mathbf{W}^t \right\|_F^2 \\ & \quad + \left\| \tilde{\mathbf{B}}^T \mathbf{X}^s \mathbf{W}^s - \tilde{\mathbf{B}}^T \mathbf{X}^{t_{au}} \right\|_F^2 + \lambda_3 \left( \left\| \mathbf{W}^s \right\|_1 + \left\| \mathbf{W}^t \right\|_1 \right). \end{aligned}$$

The TTSL model in (2) can be effectively solved by the alternating direction method (ADM). Among the solving procedures, the step of optimizing $\tilde{\mathbf{B}}_i$ is the key one, which can be updated by lots of widely used algorithms, such as iterative thresholding (IT) [9], accelerated proximal gradient (APG) [40], exact augmented Lagrange multiplier (EALM) [41], and inexact ALM (IALM) [41]. The detailed procedures for optimizing TTSL can be referred to [18].

## IV. TTRM Plus ASSM for Unsupervised Cross-Domain Microexpression Recognition

In this section, we introduce the proposed ASSM and TTRM in detail and then show how to use them to deal with unsupervised cross-domain microexpression recognition tasks.

### A. From TTSL to TTRM

To make the original TTSL be solvable, we relax it by the following three operations: 1) let $\mathbf{A}$ be an identity matrix; 2) let $\mathbf{B} = \mathbf{C}$; and 3) let $\omega_i$ be a non-negative rational number. Then, we are able to get the loss function term $f_1^{\text{TTRM}}$ and the group sparse term $f_2^{\text{TTRM}}$ of TTRM formulated as follows:

$$\begin{aligned} f_1^{\text{TTRM}}\left(\mathbf{Y}^{t_{au}}, \mathbf{C}_i, \omega_i\right) &= \left\| \left[\mathbf{Y}^s, \mathbf{Y}^{t_{au}}\right] - \sum_{i=1}^{M} \omega_i \mathbf{C}_i^T \left[\mathbf{X}_i^s, \mathbf{X}_i^{t_{au}}\right] \right\|_F^2 \\ \text{where } & \mathbf{y}_j^{t_{au}} \succeq 0, \ \mathbf{1}^T \mathbf{y}_j^{t_{au}} = 1, \ \omega_i \geq 0 \end{aligned} \tag{3}$$

and

$$f_2^{\text{TTRM}}(\omega_i) = \sum_{i=1}^{M} \omega_i, \ \text{ where } \omega_i \geq 0. \tag{4}$$

Note that $\omega_i$ with the value of 0 disables its corresponding projection matrix $\mathbf{C}_i$ in the projection procedures, which is actually similar to the group sparse matrix $\mathbf{B}_i$ learned by TTRLSR in (2). Furthermore, it is worth mentioning that replacing the binary constraint of $\omega_i$ with a non-negative one makes TTRM become more reasonable because a different facial local region usually has different contributions to distinguish microexpressions [42], and weighted parameters of TTRM can measure such contributions. In contrast, the learned group sparse projection matrix $\mathbf{B}$ of TTRLSR can only determine whether each feature group of its associated facial local region needs to be selected.

In addition, in our previous work of [43], we proposed to eliminating the distribution mismatch between the features from source and target domains by minimizing their mean vector difference and covariance matrix difference. Following this work, in this paper, we consider simply employing the distance of the projected mean vectors of the two datasets in the projected feature space to determine the optimal projection matrix. Under this consideration, the optimal projection matrix $\mathbf{C}$ and weighted parameter $\omega_i$ for the difference elimination term $f_3^{\text{TTRM}}$ of TTRM should be the ones that minimize the following objective function:

$$f_3^{\text{TTRM}}(\mathbf{C}_i, \omega_i) = \left\| \frac{1}{N_s} \sum_{i=1}^{M} \omega_i \mathbf{C}_i^T \mathbf{X}_i^s \mathbf{1}_s - \frac{1}{N_{t_{au}}} \sum_{i=1}^{M} \omega_i \mathbf{C}_i^T \mathbf{X}_i^t \mathbf{1}_{t_{au}} \right\|^2 \tag{5}$$

where $\mathbf{1}_s$ and $\mathbf{1}_{t_{au}}$ are the vectors whose elements are all 1 and the lengths are $N_s$ and $N_{t_{au}}$, respectively.

Finally, by combining the above three components, that is, (3)–(5) together, we are able to obtain the optimization problem of TTRM as follows:

$$\begin{aligned} \min_{\mathbf{Y}^{t_{au}}, \mathbf{A}, \mathbf{B}_i, \omega_i} \quad & f_1^{\text{TTRM}} + \lambda_1 f_2^{\text{TTRM}} + \lambda_2 f_3^{\text{TTRM}} \\ \text{s.t.} \quad & \omega_i \geq 0, \ \mathbf{y}_j^{t_{au}} \succeq 0, \ \mathbf{1}^T \mathbf{y}_j^{t_{au}} = 1. \end{aligned} \tag{6}$$

### B. Selecting the Satisfactory Auxiliary Set for TTRM Using ASSM

Similar to TTRLSR and TTSL, it is still an important problem for TTRM to select an optimal auxiliary set from the target domain. In this paper, we propose a simple yet effective method called ASSM based on the maximum mean discrepancy (MMD) [44] to select an optimal auxiliary set. MMD is proposed by Borgwardt *et al.* to compare distributions between two datasets in the reproducing kernel Hilbert space (RKHS), which is defined as

$$\text{MMD}(U, V) = \left\| \frac{1}{n_1} \sum_{k=1}^{n_1} \Phi(\mathbf{u}_k) - \frac{1}{n_2} \sum_{k=1}^{n_2} \Phi(\mathbf{v}_k) \right\|_{\mathcal{H}} \tag{7}$$

where $U = \{\mathbf{u}_1, \ldots, \mathbf{u}_{n_1}\}$ and $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_{n_2}\}$ are two different datasets. It had been proved that given a kernel mapping operator $\Phi$, $U$ and $V$ will have the same or similar distributions, if their MMD is minimized in such an RKHS.

Motivated by the work of MMD, we design two criteria to seek an optimal auxiliary set from the target domain. Specifically, such an ideal auxiliary set should satisfy the following two conditions, that is, given an RKHS, both: 1) MMD between the optimal selected auxiliary set and the source domain and 2) MMD between the optimal selected auxiliary set and the target domain should be minimized. In this case, the selected auxiliary set would not only describe the feature distribution of the target domain but also have the same or similar distribution to the source domain in the given RKHS. Following this idea, we endow a selection parameter $\alpha_k$ to each sample in the target domain whose value is either 1 or 0 to determine whether its corresponding sample can be served as an element of the optimal selected auxiliary set. Then, we can easily arrive at the proposed ASSM model which has the following formulation:

$$\begin{aligned}\min_{\alpha_i} \ & \left\| \frac{1}{\sum_{k=1}^{N_t} \alpha_k} \sum_{k=1}^{N_t} \alpha_i \Phi\left(\mathbf{x}_k^{\mathcal{V}_t}\right) - \frac{1}{N_t} \sum_{k=1}^{N_t} \Phi\left(\mathbf{x}_k^{\mathcal{V}_t}\right) \right\|_{\mathcal{H}}^2 \\ & + \left\| \frac{1}{\sum_{k=1}^{N_t} \alpha_k} \sum_{k=1}^{N_t} \alpha_i \Phi\left(\mathbf{x}_k^{\mathcal{V}_t}\right) - \frac{1}{N_s} \sum_{i=1}^{N_s} \Phi\left(\mathbf{x}_i^{\mathcal{V}_s}\right) \right\|_{\mathcal{H}}^2 \\ \text{s.t.} \ & \alpha_i \in \{0, 1\} \end{aligned} \tag{8}$$

where $\Phi$ is a kernel mapping operator; $\mathbf{x}_i^{\mathcal{V}_s}$ and $\mathbf{x}_i^{\mathcal{V}_t}$ are the $i$th columns of the source microexpression feature matrix $\mathbf{X}^s$ and target microexpression feature matrix $\mathbf{X}^t \in \mathbb{R}^{Md \times N_t}$, respectively; and $N_t$ is the number of the target microexpression samples.

Once the optimal selection parameter $\alpha_i$ of ASSM is obtained, the auxiliary set can be determined by choosing the samples corresponding to $\alpha_i$ with the value of 1 as its elements. Unfortunately, there is no closed-form solution for the optimization problem of ASSM in (8). Motivated by the work in [45] and [46], we introduce a new variable $\beta_i$ which is equal to $\alpha_i(\sum_{i=1}^{N_t} \alpha_i)^{-1}$ for ASSM. Then, by substituting $\beta_i$ into (8), the optimization problem of ASSM can be rewritten as

$$\begin{aligned}\min_{\beta} \ & \beta^T \tilde{\mathbf{K}}^{tt} \beta - \frac{1}{N_t} \beta^T \tilde{\mathbf{K}}^{tt} \mathbf{1}_t - \frac{1}{N_s} \beta^T \tilde{\mathbf{K}}^{ts} \mathbf{1}_s \\ \text{s.t.} \ & 0 \le \beta_i \le 1, \ \mathbf{1}^T \beta = 1 \end{aligned} \tag{9}$$

where $\tilde{\mathbf{K}}^{tt} = \Phi(\mathbf{X}^t)^T \Phi(\mathbf{X}^t)$, $\tilde{\mathbf{K}}^{ts} = \Phi(\mathbf{X}^t)^T \Phi(\mathbf{X}^s)$, and $\mathbf{1}_s$ and $\mathbf{1}_t$ are the vectors of all ones with the lengths of $N_s$ and $N_t$, respectively. Note that $\beta$ is theoretically sparse because $\alpha_i$ with value of 0 causes its corresponding $\beta_i$ to be 0. Since the sparseness of $\beta$ can be controlled by minimizing the L1-norm with respect to $\beta$ ($\|\beta\|_1 = \mathbf{1}^T \beta$), we relax ASSM in (9) by imposing $\mathbf{1}^T \beta$ on its objective function and then arrive at the following optimization problem:

$$\begin{aligned}\min_{\beta} \ & \beta^T \tilde{\mathbf{K}}^{tt} \beta + \beta^T \left( \lambda \mathbf{1}_t - \frac{\tilde{\mathbf{K}}^{tt} \mathbf{1}_t}{N_t} - \frac{\tilde{\mathbf{K}}^{ts} \mathbf{1}_s}{N_s} \right) \\ \text{s.t.} \ & 0 \le \beta_i \le 1. \end{aligned} \tag{10}$$

The optimization problem of the proposed ASSM can be thus converted to a standard quadratic programming (QP) one, and can be conveniently solved by lots of algorithms, such as the interior point method. Suppose we achieved the optimal $\beta_i$, then we are able to recover the binary solution for $\alpha_i$, according to the comparison between the value of $\beta_i$ and a preset threshold. Finally, the samples from the target domain corresponding to nonzero $\alpha_i$ compose the expected auxiliary set.

### C. Refining TTRM for Fitting ASSM

By using the proposed ASSM, we are able to select the optimal auxiliary set from the target domain. However, it should be noted that the selected auxiliary samples only meet the requirements in the RKHS induced by the kernel operation $\Phi$. In other words, in the original feature space, these auxiliary samples may not have the expectative distribution. Thus, TTRM learned in the original feature space based on this auxiliary set is possibly unsatisfactory. To conquer this defect, it is a good way to perform ASSM and TTRM in the same RKHS such that TTRM fits the auxiliary set selected by ASSM well. Following this idea, we build the TTRM in the RKHS which is induced for performing ASSM by the kernel mapping operator $\Phi$ and use the refined TTRM for the unsupervised cross-domain microexpression recognition instead of the original one. More specifically, let us first define a new kernel mapping operator $\phi$ that has the following relationship with $\Phi$ as follows:

$$\Phi(\mathbf{X}) = \left[\phi(\mathbf{X}_1)^T, \ldots, \phi(\mathbf{X}_M)^T\right]^T \tag{11}$$

where $\mathbf{X} = [\mathbf{X}_1^T, \ldots, \mathbf{X}_M^T]^T \in \mathbb{R}^{Md \times N}$.

Subsequently, we can obtain the refined TTRM by rewriting three key components of TTRM in the RKHS as follows:

$$\begin{aligned}\min_{\mathbf{Y}^{t_{au}}, \phi(\mathbf{C}_i), \omega_i} \ & f_1^{\text{TTRM}} + \lambda_1 f_2^{\text{TTRM}} + \lambda_2 f_3^{\text{TTRM}} \\ \text{s.t.} \ & \omega_i \ge 0, \ \mathbf{y}_j^{t_{au}} \succeq 0, \ \mathbf{1}^T \mathbf{y}_j^{t_{au}} = 1 \end{aligned} \tag{12}$$

where $f_1^{\text{TTRM}}$, $f_2^{\text{TTRM}}$, and $f_3^{\text{TTRM}}$ are expressed as follows:

$$\begin{aligned} & f_1^{\text{TTRM}}\left(\mathbf{L}^{t_{au}}, \phi(\mathbf{C}_i), \omega_i\right) \\ & \quad = \left\| \left[\mathbf{Y}^s, \mathbf{Y}^{t_{au}}\right] - \sum_{i=1}^{M} \omega_i \phi(\mathbf{C}_i)^T \left[\phi\left(\mathbf{X}_i^s\right), \phi\left(\mathbf{X}_i^{t_{au}}\right)\right] \right\|_F^2 \\ & \quad \text{s.t.} \quad \mathbf{y}_j^{t_{au}} \succeq 0, \ \mathbf{1}^T \mathbf{y}_j^{t_{au}} = 1 \\ & f_2^{\text{TTRM}}(\omega_i) = \sum_{i=1}^{M} \omega_i, \ \text{s.t.} \ \omega_i \ge 0 \end{aligned}$$

and

$$\begin{aligned} f_3^{\text{TTRM}}(\phi(\mathbf{C}_i), \omega_i) = \Bigg\| & \frac{1}{N_s} \sum_{i=1}^{M} \omega_i \phi(\mathbf{C}_i)^T \phi\left(\mathbf{X}_i^s\right) \mathbf{1}_s \\ & - \frac{1}{N_{t_{au}}} \sum_{i=1}^{M} \omega_i \phi(\mathbf{C}_i)^T \phi\left(\mathbf{X}_i^{t_{au}}\right) \mathbf{1}_{t_{au}} \Bigg\|^2. \end{aligned}$$

It is notable that the major goal of ASSM is to alleviate the data distribution differences between the source dataset

and the auxiliary dataset, which is actually the major target of minimizing the term of $f_3^{\text{TTRM}}$. Considering that the minimization operation of ASSM is performed before the TTRM optimization, it is reasonable to remove the term of $f_3^{\text{TTRM}}$ from (12) in order to reduce the model complexity of TTRM, which results in the following simplified TTRM model:

$$\min_{\mathbf{Y}^{t_{au}},\phi(\mathbf{C}_i),\omega_i} f_1^{\text{TTRM}} + \lambda_1 f_2^{\text{TTRM}} \quad \text{s.t. } \omega_i \geq 0,\ \mathbf{y}_j^{t_{au}} \geq 0,\ \mathbf{1}^T \mathbf{y}_j^{t_{au}} = 1. \tag{13}$$

According to the reproduced kernel theory, in the RKHS induced by $\phi$, the projection matrix $\phi(\mathbf{C}_i)$ can be reconstructed by $\phi(\mathbf{X}_i^s)$ and $\phi(\mathbf{X}_i^{t_{au}})$, that is, $\phi(\mathbf{C}_i) = [\phi(\mathbf{X}_i^s), \phi(\mathbf{X}_i^{t_{au}})]$, where $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_c] \in \mathbb{R}^{(N_s+N_{t_{au}})\times c}$ is a representation matrix. In addition, for a better reconstruction for $\phi(\mathbf{C}_i)$, we also impose an L1-norm with respect to $\mathbf{P}$, that is, $\|\mathbf{P}\|_1 = \sum_{i=1}^{c} \|\mathbf{p}_j\|_1$ to serve as the regularization term on the objective function of TTRM. Thus, we are able to reach the final version of TTRM, which is formulated as follows:

$$\min_{\mathbf{Y}^{t_{au}},\omega_i,\mathbf{P}} \left\| \left[\mathbf{Y}^s, \mathbf{Y}^{t_{au}}\right] - \mathbf{P}^T \sum_{i=1}^{M} \omega_i \mathbf{K}_i \right\|_F^2 + \lambda_1 \sum_{i=1}^{M} \omega_i + \lambda_2 \|\mathbf{P}\|_1 \quad \text{s.t. } \omega_i \geq 0,\ \mathbf{y}_j^{t_{au}} \geq 0,\ \mathbf{1}^T \mathbf{y}_j^{t_{au}} = 1 \tag{14}$$

where $\mathbf{K}_i = \begin{bmatrix} \mathbf{K}_i^{ss} & \mathbf{K}_i^{st} \\ \mathbf{K}_i^{ts} & \mathbf{K}_i^{tt} \end{bmatrix}$, $\mathbf{K}_i^{ss} = \phi(\mathbf{X}_i^s)^T \phi(\mathbf{X}_i^s)$, $\mathbf{K}_i^{st} = \phi(\mathbf{X}_i^s)^T \phi(\mathbf{X}_i^{t_{au}})$, $\mathbf{K}_i^{ts} = \phi(\mathbf{X}_i^{t_{au}})^T \phi(\mathbf{X}_i^s)$, $\mathbf{K}_i^{tt} = \phi(\mathbf{X}_i^{t_{au}})^T \phi(\mathbf{X}_i^{t_{au}})$, and $\lambda_1$ and $\lambda_2$ are the tradeoff parameters which control the sparsity of the weighted parameter vector $\omega$ and the reconstruction coefficient matrix $\mathbf{P}$, respectively.

### *D. Optimization of TTRM*

The optimization problem of TTRM in (14) can be efficiently solved by the ADM, which consists of two major steps: 1) fix $\mathbf{Y}^{t_{au}}$ and update $\omega$ and $\mathbf{P}$ and 2) fix $\omega$ and $\mathbf{P}$ and update $\mathbf{Y}^{t_{au}}$. We summarize its detailed solving procedures in Algorithm 1, where we can adopt IALM [41] to optimize $\mathbf{P}$ and adopt Liu *et al.*'s SLEP package [47] to learn $\omega$.

### *E. Microexpression Prediction in the Target Domain Based on TTRM*

Once the optimal solution of TTRM is obtained, we are able to estimate the microexpression categories of the samples from the target domain. More specifically, suppose that the learned optimal TTRM parameters are denoted by $\omega_*$ and $\mathbf{P}_*$. We first compute the microexpression label vector for a given testing feature vector $\mathbf{x}_{te}^t$ from the target microexpression domain by solving the QP problem as follows:

$$\min_{\mathbf{y}_{te}^t} \left\| \mathbf{y}_{te}^t - \mathbf{P}_*^T \sum_{i=1}^{M} (\omega_*)_i \left(\mathbf{K}_{te}^t\right)_i \right\|_F^2 \quad \text{s.t. } \mathbf{y}_{te}^t \geq 0,\ \mathbf{1}^T \mathbf{y}_{te}^t = 1 \tag{15}$$

**Algorithm 1:** Algorithm for Solving TTRM in (14)

**Input:** source microexpression feature matrix $\mathbf{X}^s$, source microexpression label matrix $\mathbf{Y}^s$, auxiliary microexpression feature matrix $\mathbf{X}^{t_{au}}$ selected by ASSM (Eq. (10)), kernel function $\phi$, and tradeoff parameters $\lambda_1$ and $\lambda_2$.

**Output:** model parameters $\mathbf{Y}_*^{t_{au}}$, $\omega_*$, and $\mathbf{P}_*$.

**Initialize:** $k = 0$, $\mathbf{Y}_k^{t_{au}}$, and $\omega_k$.

**While** the residue of the objective function $< \epsilon$ or $k = k_{max}$ **do**

1) Fix $\mathbf{Y}_k^{t_{au}}$ and update $\omega_{k+1}$ and $\mathbf{P}_{k+1}$:
   Initialize $t = 0$ and let $\omega_t = \omega_k$;
   **While** the residue of the objective function $< \epsilon$ or $t = t_{max}$ **do**
   a) Fix $\omega_t$ and update $\mathbf{P}_{t+1}$:
   $$\mathbf{P}_{t+1} = \arg\min_{\mathbf{P}} \|\tilde{\mathbf{Y}}_t - \mathbf{P}^T \hat{\mathbf{K}}_t\|_F^2 + \lambda_2 \|\mathbf{P}\|_1;$$
   where $\tilde{\mathbf{Y}}_k = [\mathbf{Y}^s, \mathbf{Y}_k^{t_{au}}]$ and $\hat{\mathbf{K}}_t = \sum_{i=1}^{M} (\omega_t)_i \mathbf{K}_i$.
   b) Fix $\mathbf{P}_{t+1}$ and update $\omega_{t+1}$:
   $$\omega_{t+1} = \arg\min_{\omega} \|\hat{\mathbf{y}}_k - \mathbf{V}\omega\|_F^2 + \lambda_1 \|\omega\|_1;$$
   where $\hat{\mathbf{y}}_k = vec(\tilde{\mathbf{Y}}_k)$, $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_M]$, $\mathbf{v}_i = vec(\mathbf{P}_{t+1}^T \mathbf{K}_i)$, and $vec(\cdot)$ means transforming a matrix to a vector by concatenating all the columns of the matrix.
   c) t = t + 1;
   **end while**
   We have $\omega_{k+1} = \omega_{t+1}$ and $\mathbf{P}_{k+1} = \mathbf{P}_{t+1}$.
2) Fix $\omega_{k+1}$ and $\mathbf{P}_{k+1}$ and update $\mathbf{Y}_{k+1}^{t_{au}}$:
   $$\mathbf{Y}_{k+1}^{t_{au}} = \arg\min_{\mathbf{Y}^{t_{au}}} \|[\mathbf{Y}^{t_{au}} - \mathbf{P}^T \sum_{i=1}^{M} (\omega_{k+1})_i \mathbf{K}_i^{t_{au}}\|_F^2 \quad \text{s.t. } \omega_i \geq 0,\ \mathbf{y}_j^{t_{au}} \geq 0,\ \mathbf{1}^T \mathbf{y}_j^{t_{au}} = 1.$$
   This is a standard QP problem, in this algorithm, we employ Interior Point method to solve it.
3) $k = k + 1$;

**end while**

and then assign a microexpression category to this testing sample by using the following criterion:

$$\text{microexpression\_category} = \arg\max_k \left\{\mathbf{y}_{te}^t(k)\right\} \tag{16}$$

where $\mathbf{y}_{te}^t(k)$ means the $k$th element of vector $\mathbf{y}_{te}^t$

## V. Experiments

### *A. Experimental Protocol*

We conduct extensive unsupervised cross-domain microexpression recognition experiments to evaluate the performance of our proposed TTRM + ASSM method. Two recent widely used spontaneous microexpression databases, that is, SMIC and CASME II, are adopted. The SMIC database is collected by Li *et al.* [48] from the University of Oulu and its samples are recorded by a high-speed (HS) camera of 100 frames/s,

TABLE II
SAMPLE STATISTICS OF CASME II AND SMIC DATABASES WITH THE SAME MICRO-EXPRESSION LABELS FOR UNSUPERVISED CROSS-DOMAIN MICRO-EXPRESSION RECOGNITION EXPERIMENTS

| Database | Positive | Negative | Surprise |
|---|---|---|---|
| CASME II | 32 | 91 | 25 |
| SMIC (HS) | 51 | 70 | 43 |
| SMIC (VIS) | 23 | 28 | 20 |
| SMIC (NIR) | 23 | 28 | 20 |

(a) (b) (c) (d)

Fig. 3. Examples of SIMC and CASME II microexpression databases. From left to right, they are (a) SMIC (HS), (b) SMIC (NIR), (c) SMIC (VIS), and (d) CASME II, respectively.

a normal visual (VIS) camera of 25 frames/s, and a near-infrared (NIR) camera, respectively. The HS set contains 164 samples from 16 subjects and is divided into three different microexpression categories, that is, *Positive*, *Negative*, and *Surprise*. The VIS and NIR subsets comprise 71 samples and are categorized into the same three microexpression classes as an HS subset, respectively. The CASME II database is built by Yan *et al.* [49] from the Institute of Psychology, Chinese Academy of Sciences. It consists of 247 microexpression samples from 26 participants and these samples belong to five microexpression classes, that is, *Happiness*, *Surprise*, *Disgust*, *Repression*, and *Others*, respectively.

We design two types of unsupervised cross-domain microexpression recognition experiments. The first one is based on either two sets of the SMIC (HS, VIS, and NIR) database, for example, HS versus VIS, and we denote this type of experiment and its six combinations by TYPE-I: Exp.1 (H $\rightarrow$ V), Exp.2 (V $\rightarrow$ H), Exp.3 (H $\rightarrow$ N), Exp.4 (N $\rightarrow$ H), Exp.5 (V $\rightarrow$ N), and Exp.6 (N $\rightarrow$ V). The other types of experiments are based on one set of SMIC (HS, VIS, and NIR) and CASME II, which consists of six combinations as well, and are denoted by TYPE-II: Exp.7 (C $\rightarrow$ H), Exp.8 (H $\rightarrow$ C), Exp.9 (C $\rightarrow$ V), Exp.10 (V $\rightarrow$ C), Exp.11 (C $\rightarrow$ N), and Exp.12 (N $\rightarrow$ C). Note that H, V, N, and C are short for SMIC (HS), SMIC (VIS), SMIC (NIR), and CASME II.

For the CASME II database, we first select the samples excluding *Others* category and relabel the *Happiness* samples with the *Positive* microexpression and the *Disgust* and *Repression* samples with the *Negative* one. Thus, the CASME II and three subsets of SMIC would share the common microexpression categorization. We list the sample distributions under the consistent labeling of all three microexpression databases for the experiments in Table II. We also give an example of the samples from the SMIC and CASME II databases, respectively, in Fig. 3 to show the difference among these three microexpression databases, where we list an image frame from the microexpression video clip.

Following the work of [17], the face images of the video clips from CASME II are cropped and transformed to 308 $\times$ 257 pixels, while for SMIC databases, we crop and then transform the images of microexpression samples into 170 $\times$ 139 pixels. As for performance evaluation metrics, we report the experimental results in terms of both weighted average recall (WAR) and unweighted average recall (UAR), which are widely used in the cross-domain speech emotion recognition research [19]. WAR is actually the normal recognition accuracy, while UAR is defined as the mean accuracy of each class divided by the number of classes without consideration of samples per class. Since most of the above microexpression databases, for example, CASME II, are class-imbalanced, which means the numbers of samples belonging to different classes have a large difference, it is better to evaluate the performance of the comparison methods in terms of both WAR and UAR.

For comparison purposes, we choose nine representative DA methods, that is, KMM [23], [24]; KLIEP [23], [25]; uLSIF [23], [26]; transfer component analysis (TCA) [50]; GFK [31]; subspace alignment (SA) [51]; STM [27], [28]; transfer kernel machine (TKL) [52]; and TSRG [17] to conduct the experiments under the same protocol as our TTRM + ASSM method. Meanwhile, we use the SVM without any DA to conduct the experiments to serve as the baseline. The microexpression features used for experiments and the tradeoff parameters of all methods are set as below.

*1) Micro-Expression Feature:* The spatiotemporal descriptor used for the experiments is uniform LBP-TOP [6]. Its neighboring radius $R$ and number of the neighboring points $P$ for LBP operator on three orthogonal planes are set to 3 and 8, respectively.

*2) Classifier:* Linear SVM with $C = 1$ is used to serve as the classifier for all the comparison DA methods. Note that to offer a fair comparison, the linear kernel is also used for all of the DA methods involving kernel functions in the experiments.

*3) Parameter Setting for DA Methods:* There are some important parameters for all of DA methods to be set such as the tradeoff the parameters $\lambda_1$ and $\lambda_2$ for TTRM, and the reduced dimension $k$ for TCA, which affects the performance of these methods. It is known that in unsupervised cross-domain microexpression recognition, the label information of the target samples is completely not provided. For this reason, the cross-validation method is not practical to determine the tradeoff parameters in the experiments. Consequently, to offer a fair comparison among all methods, in this paper, we use the widely used parameter space search strategy [17], [52]–[55] in unsupervised DA experiments for these methods and report the best results which correspond to the optimal parameters with a preset parameter space. The details of the parameter setting for these DA methods are as follows.

*KMM:* According to the suggestion of [24], its two important parameters including the upper limit of importance weight $B$ and $\epsilon$ are set to be 1000, and $\sqrt{n_{tr}} - (1/\sqrt{n_{tr}})$, where $n_{tr}$ denotes the number of training samples.

*KLIEP:* No parameter for KLIEP needs to be set.

TABLE III
EXPERIMENTAL RESULTS ON EITHER TWO SUBSETS OF SMIC (HS, VIS, AND NIR) DATABASES (TYPE-I) IN TERMS OF WAR/UAR. THE COMMON MICRO-EXPRESSIONS (3 CLASSES) ARE NEGATIVE, POSITIVE, AND SURPRISE. THE BEST RESULTS IN EACH EXPERIMENT ARE HIGHLIGHTED IN BOLD

| Method | Exp.1: H → V | Exp.2: V → H | Exp.3: H → N | Exp.4: N → H | Exp.5: V → N | Exp.6: N → V | Average |
|---|---|---|---|---|---|---|---|
| SVM | 76.06 / 79.55 | 59.76 / 62.30 | 70.42 / 69.63 | 53.05 / 57.17 | 69.01 / 69.14 | 69.01 / 69.45 | 66.22 / 67.87 |
| KMM [24], [23] | 81.69 / 81.06 | 48.17 / 48.90 | 59.15 / 61.76 | 46.34 / 50.88 | 70.42 / 70.80 | 66.20 / 65.60 | 62.00 / 63.17 |
| KLIEP [25], [23] | 78.87 / 79.55 | 59.76 / 62.30 | 69.01 / 68.44 | 53.05 / 57.35 | 70.42 / 70.80 | 67.61 / 68.00 | 66.45 / 67.74 |
| uLSIF [26], [23] | 85.92 / 86.32 | 60.98 / 63.85 | **77.46** / **77.93** | 55.49 / 59.61 | 73.24 / 73.19 | 78.87 / 78.64 | 72.00 / 73.26 |
| TCA [50] | 80.28 / 80.82 | 59.76 / 60.95 | 60.59 / 59.91 | 57.32 / 58.50 | 66.20 / 65.07 | 70.42 / 70.46 | 65.76 / 65.95 |
| GFK [31] | 83.10 / 84.76 | 59.76 / 61.58 | 69.01 / 70.00 | 62.20 / 64.82 | 74.65 / **76.32** | **83.10** / **83.68** | 71.97 / 73.53 |
| SA [51] | 84.51 / 85.65 | 62.20 / 64.69 | 71.83 / 73.51 | 54.27 / 58.08 | 70.42 / 71.06 | 80.82 / 79.59 | 70.68 / 72.10 |
| STM [27], [28] | **90.14** / 90.11 | 61.59 / 63.55 | 76.06 / 74.57 | 55.49 / 59.42 | 71.83 / 72.73 | 76.06 / 76.56 | 71.86 / 72.82 |
| TKL [52] | 74.65 / 74.99 | 62.20 / 64.21 | 69.01 / 69.53 | 56.10 / 55.55 | 73.24 / 75.13 | 73.24 / 72.49 | 68.07 / 66.98 |
| TSRG [17] | 88.73 / 88.03 | 62.80 / 64.88 | 70.42 / 71.88 | 62.20 / 63.56 | 73.24 / 74.40 | **83.10** / 83.25 | 73.42 / 74.33 |
| TTRM + ASSM | **90.14** / **90.45** | **71.34** / **73.15** | 76.06 / 75.17 | **65.24** / **66.02** | **76.06** / 73.44 | 78.87 / 79.42 | **76.29** / **76.28** |

TABLE IV
EXPERIMENTAL RESULTS ON CASME II AND THE ONE SUBSET OF SMIC (HS, VIS, AND NIR) DATABASES (TYPE-II) IN TERMS OF WAR/UAR. THE COMMON MICRO-EXPRESSIONS (3 CLASSES) ARE NEGATIVE, POSITIVE, AND SURPRISE. THE BEST RESULTS IN EACH EXPERIMENT ARE HIGHLIGHTED IN BOLD

| Method | Exp.7: C → H | Exp.8: H → C | Exp.9: C → V | Exp.10: V → C | Exp.11: C → N | Exp.12: N → C | Average |
|---|---|---|---|---|---|---|---|
| SVM | 42.07 / 37.91 | 24.32 / 35.14 | 45.07 / 42.67 | 36.49 / 48.94 | 45.07 / 43.67 | 16.22 / 31.42 | 34.87 / 39.96 |
| KMM [24], [23] | 37.20 / 34.66 | 27.70 / 39.67 | 26.76 / 25.72 | 29.73 / 44.05 | 26.76 / 23.14 | 21.62 / 36.54 | 28.30 / 33.96 |
| KLIEP [25], [23] | 45.73 / 42.32 | 24.32 / 35.14 | 45.07 / 41.93 | 36.49 / 49.91 | 47.89 / 45.58 | 16.22 / 31.71 | 33.95 / 41.10 |
| uLSIF [26], [23] | 47.56 / 47.91 | **66.22** / 43.43 | 54.93 / 53.12 | 39.19 / 53.31 | 50.70 / 51.51 | 29.05 / 40.26 | 47.94 / 48.26 |
| TCA [50] | 49.39 / 46.25 | 62.16 / 45.69 | 60.56 / 58.96 | 56.08 / 49.53 | 50.70 / 52.07 | 39.19 / 33.69 | 53.01 / 47.70 |
| GFK [31] | 51.22 / 47.75 | 62.84 / **53.19** | 59.13 / 57.21 | 49.32 / **60.65** | 43.66 / 41.61 | 21.62 / 30.63 | 47.97 / 48.51 |
| SA [51] | 48.78 / 43.16 | 60.14 / 44.68 | 45.07 / 46.64 | 48.65 / 45.10 | 45.07 / 43.53 | 49.32 / 36.19 | 49.51 / 43.22 |
| STM [27], [28] | 46.34 / 49.64 | 58.11 / 41.17 | 56.34 / 53.10 | 51.35 / 39.99 | 46.48 / 41.27 | **60.81** / 34.32 | 53.23 / 43.24 |
| TKL [52] | 51.83 / **49.77** | 56.76 / 48.65 | 47.89 / 48.17 | 45.95 / 56.20 | 43.66 / 44.25 | 28.38 / 31.59 | 45.75 / 46.44 |
| TSRG [17] | 50.61 / 47.48 | 64.19 / 49.58 | 60.56 / 58.96 | 52.03 / 56.03 | 45.07 / 45.80 | 42.57 / **43.94** | 52.51 / 50.30 |
| TTRM + ASSM | **53.05** / 45.94 | 59.46 / 50.50 | **63.38** / **63.98** | **63.51** / 47.26 | **56.34** / **57.73** | 38.51 / 41.66 | **55.71** / **51.18** |

*uLSIF:* Following the setting in [17], we search its optimal tradeoff parameter $\lambda$ from [1:1:100] $\times$ $t$ ($t = 1, 10, 100, 1000, 10\,000, 100\,000$).

*TCA, GFK, and SA:* For the experiments of these three methods, we search the optimal dimension $k$ (the number of eigenvectors for composing the projection matrix) by trying all possible dimensions, that is, searching $k \in [1, 2, \ldots, k_{\max}]$.

*STM:* As the work of [17] suggested, the searching space of the tradeoff parameter $\lambda$ for STM is set as [0.01:0.01:0.09, 0.1:0.1:1, 2:15].

*TKL:* The eigenspectrum damping factor $\zeta$ of TKL is selected by searching from the parameter space [0.1:0.1:5].

*TSRG:* TSRG has two important tradeoff parameters, that is, $\lambda$ and $\mu$. Its optimal values are determined by searching from [0.001, 0.01, 0.1, 1, 10, 100, 1000] for $\lambda$ and [0.001:0.001:0.009, 0.01:0.01:0.09, 0.1:0.1:1, 2:10] for $\mu$.

*TTRM + ASSM:* For our ASSM method, the parameter $\lambda$ is fixed at 1 throughout the experiments. As for the parameters of TTRM, the preset spaces for $\lambda_1$ and $\lambda_2$ are [0.01:0.01:0.1, 0.2:0.1:1] and [0.1:0.1:2], respectively.

### B. Experimental Results and Analysis

In this section, we report the results of all methods on the unsupervised cross-domain microexpression recognition experiments under the aforementioned protocol. The WAR and UAR of TYPE-I and TYPE-II experiments are given in Tables III and IV, respectively, where we also calculate the average among the WAR and UAR of six experiments in each type of experiment. As the experiments show, compared with the baseline results achieved by SVM without any DA, our TTRM + ASSM method has promising improvements. Furthermore, our TTRM + ASSM method also has better overall performance than all of the comparison methods, where our method achieves the highest average WAR/UAR of 76.29%/76.28% in the TYPE-I experiment and 55.71%/51.18% in the TYPE-II experiment, respectively. More specifically, the proposed TTRM + ASSM achieves the best results in terms of both WAR and UAR in most of the experiments, including Exps.1, 2, and 4 (TYPE-I), and Exps.9 and 11 (TYPE-II), and the best results in terms of WAR in several experiments, including Exp.5 (TYPE-I) and Exps.7 and 10 (TYPE-II), respectively. In addition, some comparison methods outperform the proposed TTRM + ASSM in some cases (e.g., uLSIF in Exp.3 and GFK in Exp.6). Nevertheless, we can find that the results are actually very competitive between them and our TTRM + ASSM in such cases.

Moreover, as shown in Tables III and IV, both WAR and UAR achieved by all methods in TYPE-I experiments are much higher than those in TYPE-II experiments from the comparison between the results of TYPE-I and TYPE-II. This

finding indicates that the unsupervised cross-domain microexpression recognition tasks between either of the two subsets of SMIC are easier than the tasks between one subset of SMIC and CASME II. It may be attributed to the fact that the subjects of the samples from HS, NIR, and VIS are completely the same, and these samples are just recorded by different cameras at the same time. We also notice that among the experiments of TYPE-I involving the HS dataset (Exps.1–4), the results of using the VIS dataset as the source (Exp.2) and target domain (Exp.1) are clearly at a lower level than the corresponding results of experiments involving NIR (Exps.3 and 4). We think that this is caused by the heterogeneous problem existing between the NIR dataset and HS (VIS). As the examples in Fig. 3 show, it is clear that the image quality of NIR samples is so different from that of the HS and VIS samples.

Besides, it can be from the TYPE-II experiments found that compared with the experiments of using CASME II as the source domain (Exps.7, 9, and 11), a larger gap between WAR and UAR widely exists when CASME II is served as the target domain (Exps.8, 10, and 12) for all DA methods. Based on this finding, we agree that the class imbalance problem existing in the target domain is another interference factor raising the level of difficulty on unsupervised cross-domain microexpression recognition, which has been demonstrated in [17]. As Table II shows, the CASME II database is very class-imbalanced. Its numbers of samples belonging to different microexpressions for experiments are quite different, where the largest sample number (*Negative*) is 91 while the smallest one (*Surprise*) is only 25.

### C. Comparison Between ASSM and Random Auxiliary Selection Method

Compared with our previous work of TTSL [18], one of the most worthy-mentioned advantages of this paper is that we have proposed an effective ASSM for TTRM to select a satisfactory auxiliary set instead of the random selection used in TTSL. Here, we should emphasize that the proposed ASSM is more suitable for TTRM than the random selection method. By using ASSM together, the proposed TTRM has stable performance. To check this point, we conduct additional experiments (Exps.3 and 7) by using TTRM, where its corresponding auxiliary sets are selected by ASSM and the random selection method, respectively. For ASSM, the parameter $\lambda$ is still set as 1. As for the random selection method, we randomly select the auxiliary set whose sample number is the same as ASSM and then perform each experiment (Exps.3 and 7) three times. The experimental results in terms of both WAR and UAR are given in Table V, where the results of TTRM with ASSM are directly taken from Tables III and IV. From the results, it is clear to see that the performance of TTRM with the random selection method is very unstable although it performs well in some cases, for example, Random#1 in Exp.3 and Random#2 in Exp.7. For example, it can be seen that in Exp.3, the other two results of the random selection method are both unsatisfactory, where its worst results are only 70.69%/69.98% and very similar to the baseline results (70.42%/69.63%) obtained by SVM without any DA. We think

TABLE V
COMPARISON BETWEEN ASSM AND RANDOM SELECTION METHOD, WHERE THE RESULTS ARE REPORTED IN TERMS OF WAR/UAR

| DA Method | Auxiliary Set | Exp.3 | Exp.7 |
|---|---|---|---|
| TTRM | Random#1 | 76.06 / 75.82 | 50.00 / 48.57 |
| | Random#2 | 70.69 / 69.98 | 54.27 / 51.36 |
| | Random#3 | 71.83 / 70.17 | 45.73 / 43.57 |
| | ASSM | 76.06 / 75.17 | 53.05 / 45.94 |

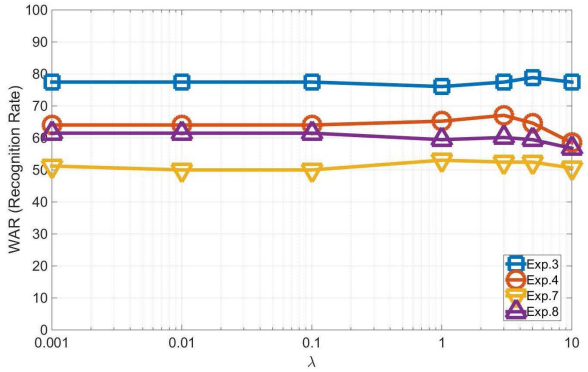


Fig. 4. Parameter sensitivity experiments for ASSM, where the results are reported in term of WAR.

the main cause of such unstabilization in the use of random selection with TTRM is that it is not satisfactory for the random selection method to remove the difference elimination term $f_3^{\text{TTRM}}$ from the objective function of the original TTRM in (12). On the other hand, it can be seen that our proposed ASSM performs both well in Exps.3 and 7, which demonstrates the advantages of the proposed ASSM over the random selection method.

### D. Evaluating ASSM With Different $\lambda$

It is clear that the sample number and the elements of the auxiliary set selected by ASSM are determined by the tradeoff parameter $\lambda$. This thus leads to an interesting question as to whether the performance of ASSM is sensitive to the selection of $\lambda$. To investigate this point, we conduct Exps.3, 4, 7, and 8, respectively, by using ASSM with different $\lambda \in \{0.001, 0.01, 0.1, 1, 3, 5, 10\}$. The WAR of ASSM with respect to the change of $\lambda$ is shown in Fig. 4. From Fig. 4, it can be seen that the performance of ASSM varies slightly with respect to the change of $\lambda$ in all experiments, which indicates that our ASSM is less sensitive to its tradeoff parameter. However, it should be pointed out that given a $\lambda$, the proposed ASSM method can only select the optimal auxiliary set under this fixed parameter. It is still unclear how to determine the optimal $\lambda$ for ASSM, which is one point of the limitations in this paper.

### E. Evaluating TTRM Plus ASSM With Different Kernel Functions

Since the proposed TTRM + ASSM is a kernel-based method, its performance is surely affected by the selection of the kernel function. Therefore, in this section, we

TABLE VI
RESULTS (WAR/UAR) OF EXPERIMENTS USING TTRM + ASSM WITH DIFFERENT KERNEL FUNCTIONS

| Kernel Function | Linear | Polynomial | ChiSquire |
|---|---|---|---|
| Exp.3 | 76.06 / 75.17 | **77.46** / **76.62** | 61.97 / 61.72 |
| Exp.7 | **53.05** / 45.94 | **53.05** / 45.88 | 48.78 / **49.94** |

choose Exp.3 (H→N) and Exp.7 (C→H) as the representatives and conduct additional experiments using TTRM + ASSM with several widely used kernel functions, including Polynomial and ChiSquare kernels. The polynomial kernel is defined as $ker(\mathbf{x}, \mathbf{y}) = (a\mathbf{x}^T\mathbf{y} + b)^c$ and in the experiments, we set its kernel parameters as $a = 1, b = 0$, and $c = 1.05$. The definition of the ChiSquare kernel function is $ker(\mathbf{x}, \mathbf{y}) = 1 - \sum_{i=1}^{d}((x_i - y_i)^2/0.5(x_i + y_i))$, where $x_i$ and $y_i$, are the $i$th elements of the feature vectors $\mathbf{x}$ and $\mathbf{y}$ whose dimension is $d$, respectively. The experimental results are shown in Table VI, where we also take the results of the linear kernel from Tables III and IV. From the results, it is interesting to see that the Polynomial kernel performs the best in terms of both WAR and UAR in Exp.3 and the ChiSquare kernel achieves the highest UAR in Exp.7. This indicates that choosing a suitable kernel benefits the performance increase of the proposed TTRM + ASSM method in dealing with the unsupervised cross-database microexpression recognition tasks. However, it is hard to determine which kernel function suits TTRM + ASSM in different tasks, which is worth a deep investigation in the future.

### *F. Convergence Analysis for the Optimization Algorithm of TTRM*

In this section, we analyze the convergence of Algorithm 1. As Algorithm 1 shows, we can see that the proposed algorithm divided the original optimization problem of TTRM in (14) into two minimization subproblems. Consequently, when the new updated variables of each minimization subproblem were obtained, the objective function value of TTRM in (14) would decrease to be smaller than the one before updating. In other words, the original objective function value of TTRM in (14) would continually decrease if we iteratively solved these two minimization subproblems. On the other hand, since the objective function of TTRM is a continuous function and lower bounded, the convergence of the proposed iterative algorithm for optimizing the TTRM problem is guaranteed. We also choose Exp.7 as the example and plot the objective function value change of TTRM with respect to the iteration of Algorithm 1, which is shown in Fig. 5. From this figure, it is clear that the objective function value of TTRM decreases and quickly reaches convergence within 10 iterations in this experiment, which indicates that the proposed algorithm for solving the TTRM optimization problem converges easily. Here, we also give the execution time of Algorithm 1 as well as the optimization problem in Section IV-E for prediction for Exp.7. By using a computer which has an Intel Core i7-4790k of 4.00 GHz and 32-GB RAM, the execution time of the algorithm and prediction is around 10.47 s and 2.71 s.

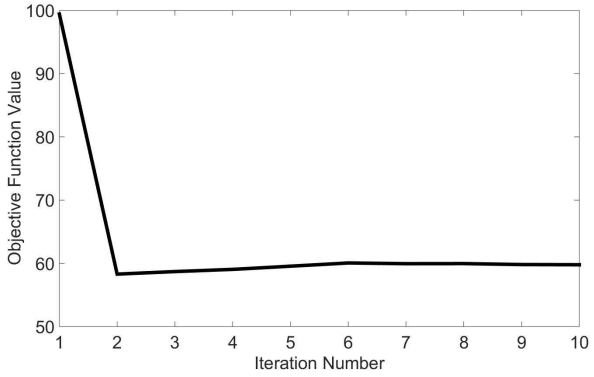


Fig. 5. Value changes of the objective function of TTRM in Exp.7 (C→H) with respect to the iteration, where the tradeoff parameters of TTRM are set as $\lambda_1 = 0.3$ and $\lambda_2 = 1.9$.

TABLE VII
STATISTICS OF USPS+MNIST AND OFFICE-10 DATASETS

| Dataset | Type | #Samples | #Features | #Classes |
|---|---|---|---|---|
| USPS | Digit | 1800 | 256 | 10 |
| MNIST | Digit | 2000 | 256 | 10 |
| Office (Amazon) | Object | 958 | 800 | 10 |
| Office (Webcam) | Object | 295 | 800 | 10 |
| Office (DSLR) | Object | 157 | 800 | 10 |

### *G. Experiments on Other Applications*

The proposed TTRM + ASSM method can be also used in many other DA applications rather than the cross-domain microexpression recognition. In this section, we would like to conduct another two types of DA experiments, that is, cross-domain handwritten digit recognition and cross-domain object recognition, to further evaluate the proposed method. To this end, two widely used benchmark datasets including USPS+MNIST[1] and Office-10 [56] are used, which results in EIGHT DA experiments: 1) USPS → MNIST; 2) MNIST → USPS; 3) A → W; 4) W → A; 5) A → D; 6) D → A; 7) W → D; and 8) D → W. Table VII shows their detailed information. For the experiments between USPS and MNIST, we follow the experimental setting of [55] and randomly sample 1800 images in USPS and 2000 images in MNIST to serve as source and target datasets alternatively. The features in this type of experiment are the gray-scale pixel values after rescaling all images to the size of $16 \times 16$. For the experiments on Office-10, we adopt the SURF features released by Gong *et al.* [31]. Five well-performing DA methods (TCA [50], GFK [31], SA [51], TKL [52], and TSRG [17]) and SVM without any DA are included in the comparison. The experimental results are given in Table VIII. From Table VIII, it is clear to see that our proposed TTRM + ASSM method achieves the best average accuracy among all methods in the experiments and outperforms all other methods in most cases (FOUR DA experiments). This shows that our method also has superior performance in dealing with the DA tasks in other applications.

[1]USPS: http://www-i6.informatik.rwth-aachen.de/~keysers/usps.html and MNIST: http://yann.lecun.com/exdb/mnist.

TABLE VIII
ACCURACY (%) FOR CROSS-DOMAIN HANDWRITTEN DIGIT RECOGNITION AND CROSS-USING TTRM + ASSM WITH DIFFERENT KERNEL FUNCTIONS

| DA Task | SVM | TCA [50] | GFK [31] | SA [51] | TKL [52] | TSRG [17] | TTRM + ASSM |
|---|---|---|---|---|---|---|---|
| USPS → MNIST | 27.30 | 40.10 | 35.70 | 42.60 | 47.80 | 34.10 | 40.70 |
| MNIST → USPS | 42.83 | 51.17 | 46.67 | 47.22 | 42.11 | 45.11 | **57.50** |
| A → W | 31.53 | 33.22 | 35.93 | 35.93 | 34.92 | 37.63 | **38.64** |
| W → A | 32.46 | 34.86 | 34.86 | 34.66 | 32.57 | 33.30 | **34.97** |
| A → D | 40.76 | 41.40 | 40.13 | 42.04 | 37.58 | **44.59** | 41.40 |
| D → A | 28.29 | 33.92 | 34.76 | **35.39** | 33.40 | 29.44 | 28.29 |
| W → D | 80.28 | 77.71 | **85.99** | **85.99** | 83.44 | 81.53 | 84.71 |
| D → W | 67.80 | 60.68 | 71.19 | 71.53 | 69.49 | 69.49 | **74.24** |
| Average | 43.90 | 46.60 | 48.15 | 49.42 | 47.66 | 46.90 | **50.06** |

## VI. CONCLUSION

In this paper, we have investigated the unsupervised cross-domain microexpression recognition problem and proposed an effective method consisting of a TTRM and an ASSM. In our method, we first make use of the proposed ASSM to select an optimal auxiliary set from the target domain for TTRM. Then, TTRM can be learned based on both the source samples and the selected auxiliary samples. By using an RKHS to perform TTRM, ASSM and TTRM can be integrated organically such that TTRM is truly suitable for the auxiliary set selected by ASSM. Extensive unsupervised cross-domain microexpression recognition experiments on the CASME II and SMIC databases are conducted to evaluate the proposed TTRM with the ASSM method. Compared with the recent state-of-the-art DA methods, our method achieves overall more promising results.

Recently, deep learning methods have been applied to the research of microexpression recognition. For instance, in the work of [57], Kim *et al.* made use of the convolutional neural network (CNN) [58] and long short-term memory (LSTM) recurrent network [59] to design a deep neural-network method to deal with the microexpression recognition problem. On the other hand, by leveraging the strong nonlinear mapping (representation) ability of deep neural networks [60], in recent years, a large number of deep transfer learning methods have been proposed and shown their promising performance in DA tasks [61]–[63]. For example, Shu *et al.* [61] proposed a weakly shared deep transfer network to deal with the cross-domain translation problem. In the work of [63], Long *et al.* designed two deep transfer networks, including the transfer deep autoencoder (TDA) and transfer deep network (TDN) for solving unsupervised DA tasks. Consequently, inspired by these successful works, we can actually develop the deep transfer learning methods to solve the unsupervised cross-domain microexpression recognition problem, which is a good direction in the future and can advance the development of this interesting and challenging topic.

## REFERENCES


[1] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, 2011, pp. 1449–1456.

[2] M. Frank, M. Herbasz, K. Sinuk, A. Keller, and C. Nolan, "I see how you feel: Training laypeople and professionals to recognize fleeting emotions," in *Proc. Annu. Meeting Int. Commun. Assoc.*, New York, NY, USA, 2009, pp. 3515–3522.

[3] M. G. Frank, C. J. Maccario, and V. Govindaraju, "Behavior and security," in *Protecting Airline Passengers in the Age of Terrorism*. Santa Barbara, CA, USA: Greenwood, 2009, pp. 86–106.

[4] M. O'Sullivan, M. G. Frank, C. M. Hurley, and J. Tiwana, "Police lie detection accuracy: The effect of lie scenario," *Law Human Behav.*, vol. 33, no. 6, pp. 530–538, 2009.

[5] H. Liang, R. Liang, M. Song, and X. He, "Coupled dictionary learning for the detail-enhanced synthesis of 3-D facial expressions," *IEEE Trans. Cybern.*, vol. 46, no. 4, pp. 890–901, Apr. 2016.

[6] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[7] J. A. Ruiz-Hernandez and M. Pietikäinen, "Encoding local binary patterns using the re-parametrization of the second order Gaussian jet," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recogn. (FG)*, Shanghai, China, 2013, pp. 1–6.

[8] S.-J. Wang, W.-J. Yan, G. Zhao, X. Fu, and C.-G. Zhou, "Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features," in *Proc. Workshop Eur. Conf. Comput. Vis.*, Zürich, Switzerland, 2014, pp. 325–338.

[9] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2009, pp. 2080–2088.

[10] X. Huang, S.-J. Wang, G. Zhao, and M. Pietikäinen, "Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Santiago, Chile, 2015, pp. 1–9.

[11] X. Huang, G. Zhao, X. Hong, W. Zheng, and M. Pietikäinen, "Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns," *Neurocomputing*, vol. 175, pp. 564–578, Jan. 2016.

[12] X. Li *et al.*, "Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 563–577, Oct./Dec. 2018.

[13] Y.-J. Liu *et al.*, "A main directional mean optical flow feature for spontaneous micro-expression recognition," *IEEE Trans. Affect. Comput.*, vol. 7, no. 4, pp. 299–310, Oct./Dec. 2016.

[14] F. Xu, J. Zhang, and J. Z. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 254–267, Apr./Jun. 2017.

[15] Z. Cui *et al.*, "Flowing on Riemannian manifold: Domain adaptation by shifting covariance," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2264–2273, Dec. 2014.

[16] M. Jiang, W. Huang, Z. Huang, and G. G. Yen, "Integration of global and local metrics for domain adaptation learning via dimensionality reduction," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 38–51, Jan. 2017.

[17] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, "Learning a target sample re-generator for cross-database micro-expression recognition," in *Proc. ACM Multimedia Conf.*, Mountain View, CA, USA, 2017, pp. 872–880.

[18] W. Zheng, Y. Zong, X. Zhou, and M. Xin, "Cross-domain color facial expression recognition using transductive transfer subspace learning," *IEEE Trans. Affect. Comput.*, vol. 9, no. 1, pp. 21–37, Jan./Mar. 2018.

[19] B. Schuller *et al.*, "Cross-corpus acoustic emotion recognition: Variances and strategies," *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, pp. 119–131, Jul./Dec. 2010.

[20] J. Deng, Z. Zhang, E. Marchi, and B. Schuller, "Sparse autoencoder-based feature transfer learning for speech emotion recognition," in *Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact. (ACII)*, Geneva, Switzerland, 2013, pp. 511–516.

[21] J. Deng, Z. Zhang, F. Eyben, and B. Schuller, "Autoencoder-based unsupervised domain adaptation for speech emotion recognition," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1068–1072, Sep. 2014.

[22] J. Deng, X. Xu, Z. Zhang, S. Frühholz, and B. Schuller, "Universum autoencoder-based domain adaptation for speech emotion recognition," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 500–504, Apr. 2017.

[23] A. Hassan, R. Damper, and M. Niranjan, "On acoustic emotion recognition: Compensating for covariate shift," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 7, pp. 1458–1468, Jul. 2013.

[24] J. Huang, A. Gretton, K. M. Borgwardt, B. Schölkopf, and A. J. Smola, "Correcting sample selection bias by unlabeled data," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 601–608.

[25] M. Sugiyama, S. Nakajima, H. Kashima, P. V. Bünau, and M. Kawanabe, "Direct importance estimation with model selection and its application to covariate shift adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2008, pp. 1433–1440.

[26] T. Kanamori, S. Hido, and M. Sugiyama, "A least-squares approach to direct importance estimation," *J. Mach. Learn. Res.*, vol. 10, pp. 1391–1445, Dec. 2009.

[27] W.-S. Chu, F. De la Torre, and J. F. Cohn, "Selective transfer machine for personalized facial action unit detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Portland, OR, USA, 2013, pp. 3515–3522.

[28] W.-S. Chu, F. De la Torre, and J. F. Cohn, "Selective transfer machine for personalized facial expression analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 529–545, Mar. 2017.

[29] E. Sangineto, G. Zen, E. Ricci, and N. Sebe, "We are not all equal: Personalizing models for facial expression analysis with transductive parameter transfer," in *Proc. 22nd ACM Int. Conf. Multimedia*, Orlando, FL, USA, 2014, pp. 357–366.

[30] K. Yan, W. Zheng, Z. Cui, and Y. Zong, "Cross-database facial expression recognition via unsupervised domain adaptive dictionary learning," in *Proc. Int. Conf. Neural Inf. Process.*, Kyoto, Japan, 2016, pp. 427–434.

[31] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2012, pp. 2066–2073.

[32] J. Zhang, Y. Han, J. Tang, Q. Hu, and J. Jiang, "Semi-supervised image-to-video adaptation for video action recognition," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 960–973, Apr. 2017.

[33] Y. Han *et al.*, "Image attribute adaptation," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 1115–1126, Jun. 2014.

[34] G. Zhao and M. Pietikäinen, "Boosted multi-resolution spatiotemporal descriptors for facial expression recognition," *Pattern Recogn. Lett.*, vol. 30, no. 12, pp. 1117–1127, 2009.

[35] T. Ojala, M. Pietikäinen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[36] W. Zheng, "Multi-view facial expression recognition based on group sparse reduced-rank regression," *IEEE Trans. Affect. Comput.*, vol. 5, no. 1, pp. 71–85, Jan./Mar. 2014.

[37] Y. Zong, W. Zheng, X. Huang, J. Yan, and T. Zhang, "Transductive transfer LDA with Riesz-based volume LBP for emotion recognition in the wild," in *Proc. ACM Int. Conf. Multimodal Interact.*, Seattle, WA, USA, 2015, pp. 491–496.

[38] Y. Zong *et al.*, "Emotion recognition in the wild via sparse transductive transfer linear discriminant analysis," *J. Multimodal User Interfaces*, vol. 10, no. 2, pp. 163–172, 2016.

[39] M. Kan, J. Wu, S. Shan, and X. Chen, "Domain adaptation for face recognition: Targetize source domain bridged by common subspace," *Int. J. Comput. Vis.*, vol. 109, nos. 1–2, pp. 94–109, 2014.

[40] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Montreal, QC, Canada, 2009, pp. 457–464.

[41] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.

[42] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, "Learning from hierarchical spatiotemporal descriptors for micro-expression recognition," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3160–3172, Nov. 2018.

[43] Y. Zong, W. Zheng, T. Zhang, and X. Huang, "Cross-corpus speech emotion recognition based on domain-adaptive least-squares regression," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 585–589, May 2016.

[44] K. M. Borgwardt *et al.*, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, 2006.

[45] B. Gong, K. Grauman, and F. Sha, "Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation," in *Proc. ICML*, Atlanta, GA, USA, 2013, pp. 222–230.

[46] L. Zhang *et al.*, "Active learning based on locally linear reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 2026–2038, Oct. 2011.

[47] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning With Efficient Projections*, vol. 6, Arizona State Univ., Tempe, AZ, USA, 2009, p. 491.

[48] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recogn. (FG)*, Shanghai, China, 2013, pp. 1–6.

[49] W.-J. Yan *et al.*, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS ONE*, vol. 9, no. 1, 2014, Art. no. e86041.

[50] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[51] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, 2013, pp. 2960–2967.

[52] M. Long, J. Wang, J. Sun, and P. S. Yu, "Domain invariant transfer kernel learning," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 6, pp. 1519–1532, Jun. 2015.

[53] Z. Ding, M. Shao, and Y. Fu, "Latent low-rank transfer subspace learning for missing modality recognition," in *Proc. AAAI*, 2014, pp. 1192–1198.

[54] M. Al-Shedivat, J. J.-Y. Wang, M. Alzahrani, J. Z. Huang, and X. Gao, "Supervised transfer sparse coding," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 1665–1672.

[55] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Columbus, OH, USA, 2014, pp. 1410–1417.

[56] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 213–226.

[57] D. H. Kim, W. J. Baddar, and Y. M. Ro, "Micro-expression recognition with expression-state constrained spatio-temporal feature representations," in *Proc. ACM Multimedia Conf.*, Amsterdam, The Netherlands, 2016, pp. 382–386.

[58] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[59] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *J. Mach. Learn. Res.*, vol. 3, pp. 115–143, Aug. 2002.

[60] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[61] X. Shu, G.-J. Qi, J. Tang, and J. Wang, "Weakly-shared deep transfer networks for heterogeneous-domain knowledge propagation," in *Proc. 23rd ACM Int. Conf. Multimedia*, Brisbane, QLD, Australia, 2015, pp. 35–44.

[62] Z. Wang *et al.*, "DeepFont: Identify your font from an image," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 451–459.

[63] M. Long, J. Wang, Y. Cao, J. Sun, and P. S. Yu, "Deep learning of transferable representation for scalable domain adaptation," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 8, pp. 2027–2040, Aug. 2016.

**Yuan Zong** (M'18) received the B.S. and M.S. degrees in electronics engineering from Nanjing Normal University, Nanjing, China, in 2011 and 2014, respectively, and the Ph.D. degree in biomedical engineering from Southeast University, Nanjing, in 2018.

He is currently a Lecturer with the Key Laboratory of Child Development and Learning Science of the Ministry of Education, School of Biological Science and Medical Engineering, Southeast University. From 2016 to 2017, he was a visiting student with the Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland. His current research interests include affective computing, pattern recognition, and computer vision.

**Wenming Zheng** (SM'18) received the B.S. degree in computer science from Fuzhou University, Fuzhou, China, in 1997, the M.S. degree in computer science from Huaqiao University, Quanzhou, China, in 2001, and the Ph.D. degree in signal processing from Southeast University, Nanjing, China, in 2004.

Since 2004, he has been with the Research Center for Learning Science, Southeast University, where he is currently a Professor with the School of Biological Science and Medical Engineering and the Key Laboratory of Child Development and Learning Science of the Ministry of Education. His current research interests include affective computing, pattern recognition, machine learning, and computer vision.

Dr. Zheng served as an Associate Editor of several peer-reviewed journals, such as the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, *Neurocomputing*, and *Visual Computer*.

**Zhen Cui** (M'14) received the B.S. degree in computer science from Shandong Normal University, Jinan, China, in 2004, the M.S. degree in computer science from Sun Yat-sen University, Guangzhou, China, in 2006, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014.

He was a Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, from 2014 to 2015. He was also a Research Assistant with Nanyang Technological University, Singapore, in 2012, for six months. He is currently a Professor with the Nanjing University of Science and Technology, Nanjing, China. His current research interests include computer vision, pattern recognition, and machine learning, especially focusing on deep learning, manifold learning, sparse coding, face detection/alignment/recognition, object tracking, image super resolution, and emotion analysis.

**Guoying Zhao** (SM'12) received the Ph.D. degree in computer science from the Chinese Academy of Sciences, Beijing, China, in 2005.

She is currently a Professor with the Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland, and a Professor with the School of Information and Technology, Northwest University, Xi'an, China. She has authored or coauthored over 190 papers in journals and conferences. She has over 9000 Google Scholar citations with an $H$-index of 43. Her current research interests include image and video descriptors, facial-expression and micro-expression recognition, dynamictexture recognition, human motion analysis, and person identification.

Dr. Zhao was the Co-Chair of many International Workshops at ECCV, ICCV, CVPR, ACCV, and BMVC. She is the Co-Publicity Chair for FG2018 and has served as the area chair for several conferences. She is currently an Associate Editor of *Pattern Recognition*, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *Image and Vision Computing Journals*.

**Bin Hu** (SM'15) received the M.S. degree in computer science from the Beijing University of Technology, Beijing, China, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Science, Beijing.

From 2007 to 2009, he was a Reader with the Head of Context Aware Computing Research Group, School of CTN, Birmingham City University, Birmingham, U.K. Since 2009, he has been the Dean of the School of Information Science and Engineering, Lanzhou University, Lanzhou, China. He is a Guest Professor of ETH Zurich, Zürich, Switzerland. His current research interests include pervasive computing, cognitive computing, and mental health care.

Dr. Hu served as an Editor for *IET Communications*, *Cluster Computing*, *Wireless Communications and Mobile Computing*, the *Journal of Internet Technology*, *Security and Communication Networks* (Wiley), and *Brain Informatics* and an Associate Editor of some peer-reviewed journals in computer science, such as the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING and the IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS. He is also an IET Fellow, an IET Fellow Assessment Panel Member (China Committee), the Co-Chair of IEEE SMC TC on *Cognitive Computing*, a Member-at-Large of ACM China, the Director of the Web Intelligence Consortium (China Committee), and a Board Member of the International Society for Social Neuroscience (China Committee).