# Seeking Salient Facial Regions for Cross-Database Micro-Expression Recognition

Xing-xun Jiang, Yuan Zong, Wen-ming Zheng
Southeast University
{jiangxingxun,xhzongyuan,wenming_zheng}@seu.edu.cn

## Abstract

*This paper focuses on the research of cross-database micro-expression recognition, in which the training and test micro-expression samples belong to different microexpression databases. Mismatched feature distributions between the training and testing micro-expression feature degrade the performance of most well-performing micro-expression methods. To deal with cross-database micro-expression recognition, we propose a novel domain adaption method called Transfer Group Sparse Regression (TGSR). TGSR learns a sparse regression matrix for selecting salient facial local regions and the corresponding relationship of the training set and test set. We evaluate our TGSR model in CASME II and SMIC databases. Experimental results show that the proposed TGSR achieves satisfactory performance and outperforms most state-of-the-art subspace learning-based domain adaption methods.*

## 1. Introduction

Micro-expression is an involuntary facial expression, which may reflect the subject's genuine emotion[5]. Discovered by Paul Ekman in 1969[6], micro-expression has played an irreplaceable role in many fields like criminal investigation[29], clinical diagnosis[8], and human-computer interaction[31, 30]. Micro-expression is a subtle facial action with tiny motion amplitude and a short duration time[6], which makes its recognition a professional task. It depends on expert psychologist and their precious time. Such a highly specialized and time-consuming task makes micro-expression recognition inefficient in practical.

To lower this barrier, researchers make some efforts to design an automatic micro-expression recognition system[23, 26, 25, 18, 13, 20, 27, 10, 38] in the last decade. Some researchers built micro-expression datasets like CASME II[28] and SMIC[16], providing the platform and benchmark to develop algorithms and methods. Some researchers make well-designed hand-crafted Spatio-temporal descriptors like LBP-TOP[33], LBP-SIP[26], Facial Dynamics Map (FDM)[27], Fuzzy Histogram of Optical Flow Orientation (FHOFO)[10] to extract micro-expression features. Some researchers[26, 18]combined micro-expression traits and cognitive processes, further promote performance. Some researchers explore automatic microexpression recognition using deep convolutional neural networks (CNNs)[14] and long short-term memory (LSTM)[12] recurrent neural networks. However, well-designed models perform far away from satisfaction because capturing microexpression's subtle movement from the different data sources (participants, acquisition equipment, and scenes) is strenuous.

To alleviate this problem and improve the robustness, researchers introduce domain adaption into automatic microexpression recognition. As an excellent way to mimic this challenging domain adaption problem, researchers proposed Cross-Database Mirco-Expression Recognition (CDMER) topic: training in one database (source) and test in the other one (target). Cross-database-related emotion recognition problems have been extensively studied in different modalities like natural image[22], speech[24], facial expression[35, 34], and EEG[36, 17]. These works provide strong baselines to the CDMER problem.

Recently, CDMER has become an active topic, and many researchers make their efforts to improve this domain adaption problem. Some researchers suppose the source and the target domain should have the same feature distribution. They used unsupervised methods to narrow the feature distribution provided by hand-crafted features without the target label. Unfortunately, it is difficult for the source domain and target under identical independent distribution (i.i.d.) in the real world. Some researchers[15, 3] noticed the established distribution difference and used semi-supervised methods to investigate improving techniques. They designed some ways to pick an auxiliary set from the target and contribute to micro-expression prediction. But it took time to select auxiliary sets[39], especially in coping with high dimensional group features of different facial regions. In general, CDMER's performance and corresponding predict speed is far away from satisfaction.

Facial region selection is an essential factor that affects emotion recognition performance. Facial region selection is aimed to select some local regions of the whole face for reducing the computational cost and improving the emotion recognition performance at the same time. Considering this, we designed a Transfer Group Sparse Regression (TGSR) to select features of efficient facial regions for the CDMER problem. We used a shared regression matrix to bridge the source and target. The projected target micro-expression features would contain discriminative information for distinguishing different micro-expressions with the shared regression matrix. We designed a joint feature distribution adaption regularization term with a learned regression matrix. This regularization ensures the original feature distribution mismatch of the source and target samples can be alleviated in label space. More importantly, we introduce a series of binary variables (either 0 or 1) and related regularization term to select compelling facial region features and improve CDMER accuracy. By setting the selected region number, we control the sparse degree of shared regression. Further, we can trade off the model accuracy and test speed. We evaluated the proposed TGSR on two widely used micro-expression databases, i.e., SMIC(HS, VIS, and NIR)[16] and CASME II[28]. The experimental results demonstrate TGSR's performance over the most recent state-of-the-art subspace learning-based domain adaptation methods for micro-expression recognition.

## 2. Method

### 2.1. The Generation of Micro-Expression Features

The first step for micro-expression recognition is extracting corresponding compelling features. As shown in Fig. 1, we used a multi-scale grid-based spatial division scheme[32] to divide a face image into $K$ blocks and extract corresponding features. Concatenating $K$ feature vector $\mathbf{x}_k$ one by one, we obtain the hierarchical features $\mathbf{x}^{\nu} = [\mathbf{x}_1^{\mathrm{T}}, \cdots, \mathbf{x}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{Kd}$ to describe a micro-expression sample. Suppose we have $N_s$ source mirco-expression samples and $N_t$ target samples. We can denote the $i$-th facial local region feature of the source micro-expression sample and the target by $\mathbf{X}_i^s \in \mathbb{R}^{d \times N_s}$ and $\mathbf{X}_i^t \in \mathbb{R}^{d \times N_t}$, respectively. And we denote feature matrices of the source by $\mathbf{X}^s = \left[\mathbf{X}_1^{s\,\mathrm{T}}, \cdots, \mathbf{X}_K^{s\,\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{Kd \times N_s}$ and the target by $\mathbf{X}^t = \left[\mathbf{X}_1^{t\,\mathrm{T}}, \cdots, \mathbf{X}_K^{t\,\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{Kd \times N_t}$, where each column is the feature vector like $\mathbf{x}^{\nu}$. Here $d$ is the feature dimension. The source labels is denoted by $\mathbf{L}_s = [\mathbf{l}_1, \cdots, \mathbf{l}_{N_s}] \in \mathbb{R}^{C \times N_s}$, where $i$-th column of $\mathbf{L}_s$, i.e., $\mathbf{l}_i = [l_{i,1}, \cdots, l_{i,C}]^{\mathrm{T}}$ is the label vector of $i$-th source micro-expression whose $j$-th element is 1 and others are all 0 if this sample conveys the $j$-th emotion.

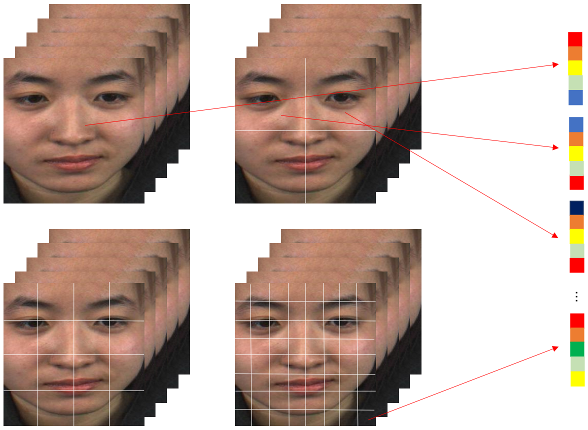

Figure 1. The multi-scale grid based spatial division scheme for feature extraction used in our experiments.

### 2.2. Building TGSR

To make the model database-independent, TGSR built a shared linear regression model to learn a regression matrix $\mathbf{C} = [\mathbf{C}_1^{\mathrm{T}}, \cdots, \mathbf{C}_K^{\mathrm{T}}] \in \mathbb{R}^{Kd \times C}$ to eliminate the feature difference between the source and target micro-expression feature, here $\mathbf{C}_i^{\mathrm{T}} \in \mathbb{R}^{d \times c}$. We first designed a regularized linear regression optimization problem for the TGSR model, which can be formulated as Equ. (1).

$$\min_{\mathbf{C_i}} \left\| \mathbf{L}^s - \sum_{i=1}^{K} \mathbf{C}_i^{\mathrm{T}} \mathbf{X}_i^s \right\|_F^2 + \xi f_1(\mathbf{C}_i) \tag{1}$$

The regularization $f_1(\mathbf{C}_i)$ is called database difference elimination one, which measures the feature difference of projected source and target features. Thus minimizing $f_1(\mathbf{C}_i)$ together with regression optimization will alleviate the large feature distribution difference. We used the maximum mean discrepancy (MMD) to serve as the regularization, which can be expressed as Equ. (2)

$$MMD\left(\mathbf{X}^s, \mathbf{X}^t\right) = \left\| \frac{1}{N_s} \sum_{i=1}^{K} \Phi\left(\mathbf{X}_i^s\right) \mathbf{1}_s - \frac{1}{N_t} \sum_{i=1}^{K} \Phi\left(\mathbf{X}^t\right) \mathbf{1}_t \right\|_{\mathcal{H}} \tag{2}$$

where $\Phi(\cdot)$ is a kernel mapping operator projecting the sameples from the original feature space to an infinite one. However, the kernel mapping operator is difficult to optimize, and we further modify the MMD into Equ. (3) to serve as $f_1(\mathbf{C}_i)$.

$$f_1\left(\mathbf{C}_i\right) = \left\| \frac{1}{N_s} \sum_{i=1}^{K} \mathbf{C}_i^{\mathrm{T}} \mathbf{X}_i^s \mathbf{1}_s - \frac{1}{N_t} \sum_{i=1}^{K} \mathbf{C}_i^{\mathrm{T}} \mathbf{X}_i^t \mathbf{1}_t \right\|_F^2 \tag{3}$$

We denote the flag of salient region as $s_i \in \{0, 1\} (i = 1, \cdots, K)$. $s_i = 1$ means that $i$-th facial local region and corresponding feature $\mathbf{X}_i$ is the salient facial local region for micro-expression recognition. Otherwise, it will not the salient if $s_i = 0$ or remove the corresponding feature $\mathbf{X}_i$. It equivalent to to replacing $\mathbf{X}_i$ with $s_i\mathbf{X}_i$ in Equ. (1). As a result, we obtain the value of $\sum_{i=1}^{K} s_i$ equal to the salient facial region number for micro-expression recognition. And the selection problem boils down to the minimization problem of $\sum_{i=1}^{K} s_i$. Combined with Equ. (3), we rewrite Equ. (1) as Equ. (4).

$$\begin{aligned}
&\min_{\mathbf{C_i}} \left\| \mathbf{L}^s - \sum_{i=1}^{K} s_i \mathbf{C}_i^{\mathrm{T}} \mathbf{X}_i^s \right\|_F^2 + \lambda \sum_{i=1}^{K} s_i \\
&+\xi \left\| \tfrac{1}{N_s} \sum_{i=1}^{K} s_i \mathbf{C}_i^{\mathrm{T}} \mathbf{X}_i^s \mathbf{1}_s - \tfrac{1}{N_t} \sum_{i=1}^{K} s_i \mathbf{C}_i^{\mathrm{T}} \mathbf{X}_i^t \mathbf{1}_t \right\|_F^2 \\
&s.t. \quad s_i \in \{0, 1\}
\end{aligned} \tag{4}$$

We note that $s_i = 0$ makes $\|s_i \mathbf{C}_i\|_F = 0$, and $\|s_i \mathbf{C}_i\|_F$ will equal to zero when $s_i = 0$ or $\mathbf{C}_i$ is the zero matrix. $\mathbf{C}_i = \mathbf{0}$ denotes corresponding facial local region $\mathbf{X}_i$ make not contribution to regression model, which is equivalent to $s_i = 0$. Consequently, by combining $s_i$ with $\mathbf{C}_i$, Equ. (4) can be rewritten as Equ. (5)

$$\begin{aligned}
&\min_{\mathbf{C}_i} \left\| \mathbf{L}^s - \sum_{i=1}^{K} \mathbf{C}_i^{\mathrm{T}} \mathbf{X}_i^s \right\|_F^2 + \lambda \sum_{i=1}^{K} \|\mathbf{C}_i\|_F \\
&+ \left\| \frac{1}{N_s} \sum_{i=1}^{K} \mathbf{C}_i^{\mathrm{T}} \mathbf{X}_i^s \mathbf{1}_s - \frac{1}{N_t} \sum_{i=1}^{K} \mathbf{C}_i^{\mathrm{T}} \mathbf{X}_i^t \mathbf{1}_t \right\|_F^2
\end{aligned} \tag{5}$$

### 2.3. Optimization of our model

In this step, the optimization problem Equ. (5) concerning $\mathbf{C}$ can be written as Equ. (6), where $\tilde{\mathbf{L}} = [\mathbf{L}^s, \mathbf{0}]$, $\tilde{\mathbf{X}}_i = \left[\mathbf{X}_i^s, \sqrt{\xi}(\frac{1}{N_s}\mathbf{X}_i^s \mathbf{1}_s - \frac{1}{N_t}\mathbf{X}_i^t \mathbf{1}_t)\right]$, $\|\mathbf{C}\|_F = \sum_{i=1}^{K} \|\mathbf{C}_i\|_F$.

$$\min_{\mathbf{C}} \left\| \tilde{\mathbf{L}} - \mathbf{C}^{\mathrm{T}} \tilde{\mathbf{X}} \right\|_F^2 + \lambda \|\mathbf{C}\|_F \tag{6}$$

We rewrite Equ. (6) as Equ. (7).

$$\begin{aligned}
&\min_{\mathbf{C},\mathbf{D}} \left\| \tilde{\mathbf{L}} - \mathbf{D}^{\mathrm{T}} \tilde{\mathbf{X}} \right\|_F^2 + \lambda \|\mathbf{C}\|_F \\
&s.t. \mathbf{D} = \mathbf{C}
\end{aligned} \tag{7}$$

Subsequently, we can obtain the corresponding augmented Lagrange function of Equ. (7) can be expressed as Equ. (8):

$$\begin{aligned}
\Gamma(\mathbf{C}, \mathbf{D}, \mathbf{P}, \mu) = &\left\| \mathbf{L} - \mathbf{D}^{\mathrm{T}} \mathbf{X} \right\|_F^2 + \lambda \|\mathbf{C}\|_F \\
&+ \mathrm{tr}\left[ \mathbf{P}^{\mathrm{T}} (\mathbf{C} - \mathbf{D}) \right] + \frac{\mu}{2} \|\mathbf{C} - \mathbf{D}\|_F^2
\end{aligned} \tag{8}$$

We used the Alternative Direction Method (ADM) and Inexact Augmented Lagrangian Multiplier method (IALM) to solve Equ. (8). We summarize the complete updating procedures in Algorithm 1.

**Algorithm 1** Algorithm of solving the optimal coefficient matrix $\mathbf{C}$ of TGSR model.

```
Input: Data matrix $\mathbf{L}$ and $\mathbf{X}$, the number of salient facial
local region $\kappa < K$ and the scalar parameter $\rho$ and $\mu_{max}$.
  • Initializing the coefficient matrix $\mathbf{C}$
  • Initializing the Largrangian multiplier matrix $\mathbf{P}$ and
    the trade off coefficient $\mu$.
Repeating steps 1) to 4) until convergence.
1: Fix $\mathbf{C}, \mathbf{P}, \mu$ and update $\mathbf{D}$:
       $\mathbf{D} = \left(\mu\mathbf{I} + 2\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)^{-1} \left(2\mathbf{X}\mathbf{L}^{\mathrm{T}} + \mathbf{P} + \mu\mathbf{C}\right)$
2: Fix other parameters and update $\mathbf{C}$
       Calculate $d_i = \left\| \mathbf{D}_i - \frac{\mathbf{P}_i}{\mu} \right\|_F$, and sort the value of
   $d_i$, such that $d_{i_1} \geq d_{i_2} \geq \cdots \geq d_{i_K}$, Let $\lambda = \mu d_{i_{\kappa+1}}$,
   then update $\mathbf{C}$.
       $$\mathbf{C}_i = \begin{cases} \frac{d_i - \frac{\lambda}{\mu}}{d_i}(\mathbf{D}_i - \frac{\mathbf{P}_i}{\mu}), & \frac{\lambda}{\mu} < d_i \\ \mathbf{0}, & \frac{\lambda}{\mu} \geq d_i \end{cases} \tag{9}$$
3: Update $\mathbf{P}$ and the regularized parameter $\mu$
       $\mathbf{P} = \mathbf{P} + \mu(\mathbf{Z} - \mathbf{C})$
       $\mu = \min(\rho\mu, \mu_{max})$
4: Check convergence:
       $\|\mathbf{C} - \mathbf{D}\|_\infty < \varepsilon$
Output:the coefficient matrix $\mathbf{C}$, the coefficient $\lambda$
```

### 2.4. TGSR for cross-database micro-expression recognition

Once our TGSR model learned the optimal parameters based on the labeled source and unlabeled target micro-expression sample, we can predict the testing micro-expression categories. Given a testing target micro-expression sample $\mathbf{x}_i^{te}$, we can estimate its micro-expression label $\mathbf{l}^{te}$ by solving the optimization problem as Equ. (10) shown.

$$\begin{aligned}
&\min_{\mathbf{l}^{te}} \left\| \mathbf{l}^{te} - \sum_{i=1}^{K} \hat{\mathbf{C}}_i^{\mathrm{T}} \mathbf{x}_i^{te} \right\|_F^2 \\
&s.t. \quad \mathbf{l}^{te} \geq 0, \mathbf{1}^{\mathrm{T}} \mathbf{l}^{te} = 1
\end{aligned} \tag{10}$$

$$\hat{c} = \arg\max_{j} \left\{ \mathbf{l}_j^{te} \right\} \tag{11}$$

where $\hat{\mathbf{C}}_i$ is the learned regression matrix of $i$-th facial local region. Finally, the micro-expression category of this testing sample is assigned as the one whose corresponding value in $\mathbf{l}^{te}$ is maximum.

Table 1. The sample information of the Selected CASME II and SMIC (HS, VIS, NIR).

| Dataset | Positive | Negative | Surprise |
|---|---|---|---|
| Selected CASME II | 32 | 73 | 25 |
| SMIC(HS) | 51 | 70 | 43 |
| SMIC(VIS) | 23 | 28 | 20 |
| SMIC(NIR) | 23 | 28 | 20 |

# 3. Experiment

## 3.1. Experiment Setup

**Database.** We evaluated our method on Selected CASME II and SMIC database. SMIC[16] is a micro-expression database that contains three subsets (HS, VIS, and NIR) from 16 subjects. HS subset is captured by high-speed cameras with 100 frames/s, while general visual cameras obtain VIS subset with 25 frames/s, and near-infrared cameras record the NIR subset. The HS subset, VIS subset, and NIR subset have 164 samples, 71 samples, and 71 samples. SMIC includes three kinds of microexpression: positive, negative, and surprise. CASME II[28] contains 247 samples from 26 participants, including five types of micro-expression, i.e., happiness, surprise, disgust, repression, and others. Selected CASME comprises 130 samples from 5 kinds of micro-expression, selecting from CASMEII database only excludes Others category and relabeling happiness into the positive class, disgust and repression into the negative category. Tab. 1 summarized the basic information of Selected CASME II and SMIC.

**Protocol.** Following [32], we employed two types of unsupervised cross-database micro-expression recognition experiments. The first one is based on either the SMIC (HS, VIS, and NIR) database, such as HS v.s. VIS. The other is based on the Selected CASME II database and either SMIC (HS, VIS, NIR), such as CASME II v.s. HS. We used $H$, $V$, $N$, and $C$ short for SMIC(HS), SMIC(VIS), SMIC(NIR), and CASME II, respectively. We can denote cross-database task from Source database $S$ to the target database $T$ by $S \to T$. For instance, a cross-database task from SMIC(HS) to CASME II can be denoted by $H \to C$.

**Evaluation Metrics.** We employed the mean F1 score and the accuracy to estimate our method. Mean F1 score can be calucated by $MeanF1 - Score = \frac{1}{c}\sum_{i=1}^{c}\frac{2p_ir_i}{p_i+r_i}$, where $p_i$ and $r_i$ are the precision and recall of the $i$-th micro-expression respectively, $c$ is the class number of micro-expressions. Mean F1 score is suitable because sample unbalanced widely existed in cross-database micro-expression recognition.

**Implemention Detail.** We used the image sequence pre-processed by the collectors of CASME II and SMIC. We employ the temporal interpolation model (TIM) to normalize the image sequences into 16 frames and resize them to $112 \times 112$. Four spatial grids $(1 \times 1, 2 \times 2, 4 \times 4, 8 \times 8)$ based spatiotemporal descriptors are served as the micro-expression features, which including $K = 85$ blocks. Then LBP-TOP[33] feature$(R = 3, P = 8)$ was extracted from each facial local region to make the multi-scale feature. We used a grid searching strategy to search the optimal hyper-parameter for our method.

## 3.2. Results and Analysis

Tab. 2 and Tab. 3 show the TYPE I and TYPE II experiment results, respectively. We calculated different methods' average performance(Mean F1-Score and WAR) of these two types of experiments. In terms of Mean F1-Score/Accuracy, SVM[2], IW-SVM[11], TCA[21], GFK[9], SA[7], STM[4], TKL[19], TcSRG[37], DRLS[40], and our method respectively achieved 0.6003/61.62, 0.6911/68.07, 0.6238/64.01, 0.7223/72.44, 0.6917/69.44, 0.6440/65.78, 0.6964/69.92, 0.6991/70.05, 0.6552/66.60 and 0.7141/71.80 in TYPE I task. And they achieved 0.4112/45.55, 0.4876/50.73, 0.5177/53.45, 0.5161/54.31, 0.4877/50.90, 0.3982/45.61, 0.4925/51.93, 0.5347/56.22, 0.5102/53.71 and 0.5424/56.33 in TYPE II task respectively. From the bold-lighted in Tab. 2 and Tab. 3, we can see that our method beat those state-of-the-art methods beyond half tasks and achieved the best performance in 7 of the total 12 tasks.

In addition, at least three apparent characteristics we can find in the tables. Firstly, We found that domain adaption tricks can promote method performance. IW-SVM vastly outperforms SVM in average performance by learning a group of importance weights: they improved 0.0911/6.45 (Mean F1-score/Accuracy) in TYPE I task and 0.0764/5.18 (Mean F1-score/Accuracy) in TYPE II task. Domain adaption-based method performs better than non-domain adaption. All methods except SVM used domain adaption tricks, and we see that they better than SVM in average performance and almost tasks steadily.

Secondly, From Tab. 2 and Tab. 3, we found that results in TYPE I generally better than TYPE II. The task itself causes it. Experiments in TYPE I selected two different modalities in the SMIC dataset, and TYPE II selected two different datasets: CASME II and one subset of the SMIC dataset. Thirdly, we also found that the exchange of the source and target have a significant performance gap. For instance, SVM's performance of Exp.1(H→V) vastly outperforms that of Exp.2(V→H); Exp.3(H→N) vastly outperforms that of Exp.4(N→H); the performance of Exp.11(C→N) vastly outperforms that of Exp.12(N→C); Exp.1(H→V) used high-speed camera captured image se-

Table 2. Experimental results (Mean F1-Score / WAR) on either two subsets of SMIC(HS, VIS and NIR) databases in terms of accuracy and meanF1. The common micro-expression(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) |
|---|---|---|---|---|---|---|
| SVM[2] | 0.8002 / 80.28 | 0.5421 / 54.27 | 0.5455 / 53.52 | 0.4878 / 54.88 | 0.6186 / 63.38 | 0.6078 / 63.38 |
| IW-SVM[11] | 0.8868 / 88.73 | 0.5852 / 58.54 | **0.7469 / 74.65** | 0.5427 / 54.27 | 0.6620 / 69.01 | 0.7228 / 73.24 |
| TCA[21] | 0.8269 / 83.10 | 0.5477 / 54.88 | 0.5828 / 59.15 | 0.5443 / 57.32 | 0.5810 / 61.97 | 0.6598 / 67.61 |
| GFK[9] | 0.8448 / 84.51 | 0.5957 / 59.15 | 0.6977 / 70.42 | 0.6197 / 62.80 | **0.7619 / 76.06** | 0.8142 / 81.69 |
| SA[7] | 0.8037 / 80.28 | 0.5955 / 59.15 | 0.7465 / **74.65** | 0.5644 / 56.10 | 0.7004 / 71.83 | 0.7394 / 74.65 |
| STM[4] | 0.8253 / 83.10 | 0.5059 / 51.22 | 0.6628 / 66.20 | 0.5351 / 56.10 | 0.6427 / 67.61 | 0.6922 / 70.42 |
| TKL[19] | 0.7742 / 77.46 | 0.5738 / 57.32 | 0.7051 / 70.42 | 0.6116 / 62.20 | 0.7558 / 76.06 | 0.7580 / 76.06 |
| TSRG[37] | 0.8869 / 88.73 | 0.5652 / 56.71 | 0.6484 / 64.79 | 0.5770 / 57.93 | 0.7056 / 70.42 | 0.8116 / 81.69 |
| DRLS[40] | 0.8604 / 85.92 | 0.6120 / 60.98 | 0.6599 / 66.20 | 0.5599 / 55.49 | 0.6620 / 69.01 | 0.5771 / 61.97 |
| Ours | **0.9150 / 91.55** | **0.6226 / 62.20** | 0.5847 / 60.56 | **0.6234 / 61.59** | 0.6984 / 70.42 | **0.8403 / 84.51** |

Table 3. Experimental results (Mean F1-Score / WAR) on CASME II and the one subset of SMIC(HS, VIS and NIR) databases in terms of accuracy and meanF1. The common micro-expression (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) |
|---|---|---|---|---|---|---|
| SVM[2] | 0.3697 / 45.12 | 0.3245 / 48.46 | 0.4701 / 50.70 | 0.5367 / 53.08 | 0.5295 / 52.11 | 0.2368 / 23.85 |
| IW-SVM[11] | 0.3541 / 41.46 | 0.5829 / 62.31 | 0.5778 / 59.15 | 0.5537 / 54.62 | 0.5117 / 50.70 | 0.3456 / 36.15 |
| TCA[21] | 0.4637 / 46.34 | 0.4870 / 53.08 | **0.6834 / 69.01** | 0.5789 / 59.23 | 0.4992 / 50.70 | 0.3937 / 42.31 |
| GFK[9] | 0.4126 / 46.95 | 0.4776 / 50.77 | 0.6361 / 66.20 | 0.6056 / 61.50 | 0.5180 / 53.52 | 0.4469 / 46.92 |
| SA[7] | 0.4302 / 47.56 | 0.5447 / 62.31 | 0.5939 / 59.15 | 0.5243 / 51.54 | 0.4738 / 47.89 | 0.3592 / 36.92 |
| STM[4] | 0.3640 / 43.90 | **0.6115 / 63.85** | 0.4051 / 52.11 | 0.2715 / 30.00 | 0.3523 / 42.25 | 0.3850 / 41.54 |
| TKL[19] | 0.4582 / 46.95 | 0.4661 / 54.62 | 0.6042 / 60.56 | 0.5378 / 53.08 | 0.5392 / 54.93 | 0.4248 / 43.85 |
| TSRG[37] | **0.5042** / 51.83 | 0.5171 / 60.77 | 0.5935 / 59.15 | 0.6208 / 63.08 | 0.5624 / 56.34 | 0.4105 / 46.15 |
| DRLS[40] | 0.4924 / **53.05** | 0.5267 / 59.23 | 0.5757 / 57.75 | 0.5942 / 60.00 | 0.4885 / 49.83 | 0.3838 / 42.37 |
| Ours | 0.5001 / 51.83 | 0.5061 / 56.92 | 0.5906 / 59.15 | **0.6403 / 63.85** | **0.5697 / 57.75** | **0.4474 / 48.46** |

quences as the source and visual camera captured as the target, which exchanged in Exp.2(V→H). We infer that a high-speed camera may capture more subtle spatial-temporal facial action, contributing to micro-expression recognition. Exp.3(N→H) used high-speed camera captured image sequences as the source and near-infrared captured image sequence as the target; the heads exchanged in Exp.4(N→H). The high-speed camera captured subset HS, and the near-infrared NIR captured are both from the SMIC dataset, but the near-infrared subset NIR discarded color information while the high-speed subset HS preserved. We found that learning micro-expression from colored to uncolored shows better performance than that from undyed to colored. It confirms that human's understanding of emotions is color-related in [1]. The same phenomenon existed in Exp.11(C→N) and Exp.12(N→C). They involved two different datasets, CASME II and SMIC. We found that the gap between Exp.11(C→N) and Exp.12(N→C) is more significant than Exp.3(N→H) and Exp.4(N→H). It tells us the dataset difference larger than the color difference.

# 4. Conclusion

In this paper, we proposed a Transfer Group Sparse Regression (TGSR) for cross-database micro-expression recognition. TGSR seeks the salient facial local region and eliminates the redundancy, thus learned a sparse project matrix from feature to label, which improves performance and speed. Furthermore, TGSR can also eliminate the discrepancies between microexpression databases effectively. Experiments show that our method surpasses most state-of-the-art domain adoption methods for micro-expression recognition.

## References

[1] Carlos F Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M Martinez. Facial color is an efficient mechanism to visually transmit emotion. *Proceedings of the National Academy of Sciences*, 115(14):3581–3586, 2018. 5

[2] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27, 2011. 4, 5

[3] Xiuzhen Chen, Xiaoyan Zhou, Cheng Lu, Yuan Zong, Wenming Zheng, and Chuangao Tang. Target-adapted subspace learning for cross-corpus speech emotion recognition. *IEICE TRANSACTIONS on Information and Systems*, 102(12):2632–2636, 2019. 1

[4] Wen-Sheng Chu, Fernando De la Torre, and Jeffery F Cohn. Selective transfer machine for personalized facial action unit detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3515–3522, 2013. 4, 5

[5] Paul Ekman. *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*. WW Norton & Company, 2009. 1

[6] Paul Ekman and Wallace V Friesen. Nonverbal leakage and clues to deception. *Psychiatry*, 32(1):88–106, 1969. 1

[7] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE international conference on computer vision*, pages 2960–2967, 2013. 4, 5

[8] Mark Frank, Malgorzata Herbasz, Kang Sinuk, A Keller, and Courtney Nolan. I see how you feel: Training laypeople and professionals to recognize fleeting emotions. In *The Annual Meeting of the International Communication Association. Sheraton New York, New York City*, pages 1–35, 2009. 1

[9] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2066–2073. IEEE, 2012. 4, 5

[10] SL Happy and Aurobinda Routray. Fuzzy histogram of optical flow orientations for micro-expression recognition. *IEEE Transactions on Affective Computing*, 10(3):394–406, 2017. 1

[11] Ali Hassan, Robert Damper, and Mahesan Niranjan. On acoustic emotion recognition: compensating for covariate shift. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(7):1458–1468, 2013. 4, 5

[12] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. 1

[13] Dae Hoe Kim, Wissam J Baddar, and Yong Man Ro. Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 382–386, 2016. 1

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012. 1

[15] Lingyan Li, Xiaoyan Zhou, Yuan Zong, Wenming Zheng, Xiuzhen Chen, Jingang Shi, and Peng Song. Unsupervised cross-database micro-expression recognition using target-adapted least-squares regression. *IEICE Transactions on Information and Systems*, 102(7):1417–1421, 2019. 1

[16] Xiaobai Li, Tomas Pfister, Xiaohua Huang, Guoying Zhao, and Matti Pietikäinen. A spontaneous micro-expression database: Inducement, collection and baseline. In *2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (fg)*, pages 1–6. IEEE, 2013. 1, 2, 4

[17] Yang Li, Wenming Zheng, Yuan Zong, Zhen Cui, Tong Zhang, and Xiaoyan Zhou. A bi-hemisphere domain adversarial neural network model for eeg emotion recognition. *IEEE Transactions on Affective Computing*, 2018. 1

[18] Yong-Jin Liu, Jin-Kai Zhang, Wen-Jing Yan, Su-Jing Wang, Guoying Zhao, and Xiaolan Fu. A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing*, 7(4):299–310, 2015. 1

[19] Mingsheng Long, Jianmin Wang, Jiaguang Sun, and S Yu Philip. Domain invariant transfer kernel learning. *IEEE Transactions on Knowledge and Data Engineering*, 27(6):1519–1532, 2014. 4, 5

[20] Ping Lu, Wenming Zheng, Ziyan Wang, Qiang Li, Yuan Zong, Minghai Xin, and Lenan Wu. Micro-expression recognition by regression model and group sparse spatio-temporal feature learning. *IEICE TRANSACTIONS on Information and Systems*, 99(6):1694–1697, 2016. 1

[21] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2010. 4, 5

[22] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009. 1

[23] Tomas Pfister, Xiaobai Li, Guoying Zhao, and Matti Pietikäinen. Recognising spontaneous facial micro-expressions. In *2011 international conference on computer vision*, pages 1449–1456. IEEE, 2011. 1

[24] Bjorn Schuller, Bogdan Vlasenko, Florian Eyben, Martin Wöllmer, Andre Stuhlsatz, Andreas Wendemuth, and Gerhard Rigoll. Cross-corpus acoustic emotion recognition: Variances and strategies. *IEEE Transactions on Affective Computing*, 1(2):119–131, 2010. 1

[25] Su-Jing Wang, Wen-Jing Yan, Xiaobai Li, Guoying Zhao, Chun-Guang Zhou, Xiaolan Fu, Minghao Yang, and Jianhua Tao. Micro-expression recognition using color spaces. *IEEE Transactions on Image Processing*, 24(12):6034–6047, 2015. 1

[26] Yandan Wang, John See, Raphael C-W Phan, and Yee-Hui Oh. Lbp with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition. In *Asian conference on computer vision*, pages 525–537. Springer, 2014. 1

[27] Feng Xu, Junping Zhang, and James Z Wang. Microexpression identification and categorization using a facial dynamics

map. *IEEE Transactions on Affective Computing*, 8(2):254–267, 2017. 1

[28] Wen-Jing Yan, Xiaobai Li, Su-Jing Wang, Guoying Zhao, Yong-Jin Liu, Yu-Hsin Chen, and Xiaolan Fu. Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. *PloS one*, 9(1):e86041, 2014. 1, 2, 4

[29] Wen-Jing Yan, Qi Wu, Jing Liang, Yu-Hsin Chen, and Xiaolan Fu. How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior*, 37(4):217–230, 2013. 1

[30] Chenguang Yang, Jing Luo, Yongping Pan, Zhi Liu, and Chun-Yi Su. Personalized variable gain control with tremor attenuation for robot teleoperation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(10):1759–1770, 2017. 1

[31] Chenguang Yang, Huaiwei Wu, Zhijun Li, Wei He, Ning Wang, and Chun-Yi Su. Mind control of a robotic arm with visual fusion technology. *IEEE Transactions on Industrial Informatics*, 14(9):3822–3830, 2017. 1

[32] Tong Zhang, Yuan Zong, Wenming Zheng, CL Philip Chen, Xiaopeng Hong, Chuangao Tang, Zhen Cui, and Guoying Zhao. Cross-database micro-expression recognition: A benchmark. *IEEE Transactions on Knowledge and Data Engineering*, 2020. 2, 4

[33] Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):915–928, 2007. 1, 4

[34] Sicheng Zhao, Chuang Lin, Pengfei Xu, Sendong Zhao, Yuchen Guo, Ravi Krishna, Guiguang Ding, and Kurt Keutzer. Cycleemotiongan: Emotional semantic consistency preserved cyclegan for adapting image emotions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2620–2627, 2019. 1

[35] Sicheng Zhao, Xin Zhao, Guiguang Ding, and Kurt Keutzer. Emotiongan: Unsupervised domain adaptation for learning discrete probability distributions of image emotions. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 1319–1327, 2018. 1

[36] Wei-Long Zheng and Bao-Liang Lu. Personalizing eeg-based affective models with transfer learning. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence*, pages 2732–2738, 2016. 1

[37] Yuan Zong, Xiaohua Huang, Wenming Zheng, Zhen Cui, and Guoying Zhao. Learning a target sample re-generator for cross-database micro-expression recognition. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 872–880, 2017. 4, 5

[38] Yuan Zong, Xiaohua Huang, Wenming Zheng, Zhen Cui, and Guoying Zhao. Learning from hierarchical spatiotemporal descriptors for micro-expression recognition. *IEEE Transactions on Multimedia*, 20(11):3160–3172, 2018. 1

[39] Yuan Zong, Wenming Zheng, Zhen Cui, Guoying Zhao, and Bin Hu. Toward bridging microexpressions from different domains. *IEEE transactions on cybernetics*, 50(12):5047–5060, 2019. 1

[40] Yuan Zong, Wenming Zheng, Xiaohua Huang, Jingang Shi, Zhen Cui, and Guoying Zhao. Domain regeneration for cross-database micro-expression recognition. *IEEE Transactions on Image Processing*, 27(5):2484–2498, 2018. 4, 5