

Cross-Dataset Facial Expression Recognition based on Arousal-Valence Emotion Model and Transfer Learning Method

Yong Yang ^a, Chuan Liu ^b and Qingshan Wu ^c

Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

^a yangyong@cqupt.edu.cn, ^b theonlylc@163.com, ^c wuqs_90@sina.com

Keywords: Facial expression recognition; Arousal-valence emotion dimensions; TPCA; Fusion.

Abstract. Traditional facial expression recognition methods assume that facial expression in the training and testing sets are collected under the same condition such that they are independent and identically distributed. However, the assumption is not satisfied in many real applications. This problem is referred to as cross-dataset facial expression recognition. On the other hand, the traditional facial expression recognition methods are based on basic emotion theory proposed by Ekman. Unfortunately, the theory is limited to express diverse and subtle emotion. To solve the problem of the cross-dataset facial expression recognition and enrich the emotion expression, a transfer learning algorithm TPCA and arousal-valence emotional model are adopted in this paper. A new facial emotion recognition method based on TPCA and two-level fusion is proposed, which combine weight fusion and correlation fusion between arousal and valence to improve the recognition performance under cross-dataset scenarios. The contrast experimental results show that the proposed method can get better recognition result than the traditional methods.

1. Introduction

Over the past decades, facial expression recognition has been widely studied in the area of computer vision and pattern recognition. Psychologist Mehrabia defined the well-known 7%-38%-55% criterion for emotion expression, that is, emotion is expressed by 7% language, 38% voice and 55% facial expression [1]. Thus facial expression is the main way to express emotion.

Basic emotion theory proposed by Ekman included six categories, such as anger, happy, sad, surprise, disgust and fear [2]. However, the theory has its limitations. Firstly, there is a limited ability to express emotions. Emoticons other than the basic expression can't be effectively expressed. Secondly, it can't express the relationship between the different emotions. Thirdly, it can't express the intensity of expression.

To solve the problem of basic emotions theory, many foreign researchers in the field of emotional computing resort to the continuous emotional model [3]. The continuous common emotional models include valence and arousal emotion model [4] proposed by Russell, pleasure, arousal, power emotional model [5] proposed Davis, the pleasure, arousal, dominance emotional dimension model [6] proposed by Mehrabian and so on. Among these models, valence and arousal model has been widely used, because it can reflect a person's emotional change most.

In arousal-valence (A-V) emotional dimension space, arousal represents the activation, and valence represents the degree of pleasantness. Both of the range is between -1 and 1. Positive correlation between A and V dimension has been verified [7]. Therefore, the positive correlation is employed in facial expression recognition research.

Recently, many machine learning methods have been used in pattern recognition successfully. In this trend, subspace learning methods have been widely used in facial expression recognition. Features of facial expression images are projected into a low dimensional feature space to reduce the feature dimensions based on these methods. Principal component analysis (PCA) [8], locality preserving projections (LPP) [9, 10] and so on are commonly used subspace methods. All conventional facial expression recognition methods assume facial expression in the training and testing sets are collected

under the same condition such that they are independent and identically distributed, however, the conditions are usually unsatisfied in many real world applications.

Performance of traditional subspace learning algorithms would be decreased significantly in the cross-dataset expression recognition because of the limited generalization ability. A new transfer subspace learning approach has been proposed to learn a feature space which transfers the knowledge gained from the training set to the testing data to improve the recognition performance under cross-dataset scenarios [11]. Experimental results have confirmed the efficacy of the proposed approach. To address the problem of facial expression recognition in cross-dataset, transfer PCA (TPCA) algorithm is used in this research.

The main contributions of this paper is that the transfer PCA is used to solve the problem of cross-dataset facial expression recognition in A-V emotion dimension model and a two-level fusion expression recognition method is presented by using the weighted fusion and A-V correlation fusion.

2. Transfer Subspace Learning Approach

2.1 Traditional Subspace Learning PCA

Let $X = \{X_1, X_2, \dots, X_N\}$, $X_i = (x_{i1}, x_{i2}, \dots, x_{iD}) \in R^D$, $i = 1, 2, \dots, N$, be a training set of facial images, where N is the number of samples and D is the feature dimension of each sample. The class label of X_i is assumed to be $L = \{l_1, l_2, \dots, l_m\}$, where m is the number of class. For the j th class, n_j denotes the number of its samples, where $j = 1, 2, \dots, m$. Hence, $N = \sum_{j=1}^m n_j$. The objective of PCA is to find a linear projection matrix $W = \{w_1, w_2, \dots, w_p\}$ to map X into a low dimensional representation Y , where $Y = W^T X \in R^p$, $p < D$, $Y = \{Y_1, Y_2, \dots, Y_N\}$, $Y_i = (y_{i1}, y_{i2}, \dots, y_{ip})$, p is the low dimensional feature dimension. PCA objective functions and constraints is

$$\min F(W) = -tr(W^T S W) \quad (1)$$

$$\text{Subject to } C(W) = W^T W - I = 0$$

$$\text{Where } S = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{m})(X_i - \bar{m})^T, \quad \bar{m} = \frac{1}{N} \sum_{i=1}^N X_i.$$

2.2 TPCA

Given N_s training samples $X_s = \{x_1^S, x_2^S, \dots, x_{N_s}^S\}$ and N_t testing samples $X_T = \{x_1^T, x_2^T, \dots, x_{N_t}^T\}$, our objective is to seek a feature space W to minimize the difference between the training and testing sets besides optimizing $F(W)$. TPCA objective function is

$$\min_W F(W) + \lambda H(W) \quad (2)$$

Where $\lambda \geq 0$, and

$$H(W) = \sum_{i=1}^{N_s} \left\| W^T x_i^S - W^T \sum_{j=1}^k r_{ij} x_{ij}^T \right\|^2 \quad (3)$$

$x_{i1}^T, x_{i2}^T, \dots, x_{ik}^T$ are the k -nearest neighbors of x_i^S , $r_{i1}, r_{i2}, \dots, r_{ik}$ are the corresponding coefficients, and they can be obtained similarly to the coefficients obtained in the locally linear embedding (LLE) method in [12]. We simplify $H(W)$ to the following form,

$$\begin{aligned} H(W) &= tr \left[W^T \sum_{i=1}^{N_s} (x_i^S - \sum_{j=1}^k r_{ij} x_{ij}^T) (x_i^S - \sum_{j=1}^k r_{ij} x_{ij}^T)^T W \right] \\ &= tr(W^T Z W) \end{aligned} \quad (4)$$

$$\text{Where } Z = \sum_{i=1}^{N_s} \left[(x_i^S - \sum_{j=1}^k r_{ij} x_{ij}^T) (x_i^S - \sum_{j=1}^k r_{ij} x_{ij}^T)^T \right].$$

To solve the minimization problem with respect to W well-posed, we utilize an orthogonal constraint $W^T W = I$ and formulate TPCA as the following constrained optimization problem,

$$\min_W T(W) = tr(W^T U W) + \lambda tr(W^T Z W) \quad (5)$$

$s.t. W^T W = I$
Let $\frac{\partial r(W)}{\partial W} = 0$, we can obtain the projections of TPCA by solving the following eigenvalue

equation:

$$(\lambda Z - U)w = \alpha w \tag{6}$$

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

3. Expression Recognition Based on TPCA

The framework of the proposed two-level fusion expression recognition approach based on TPCA is shown in Figure 1.

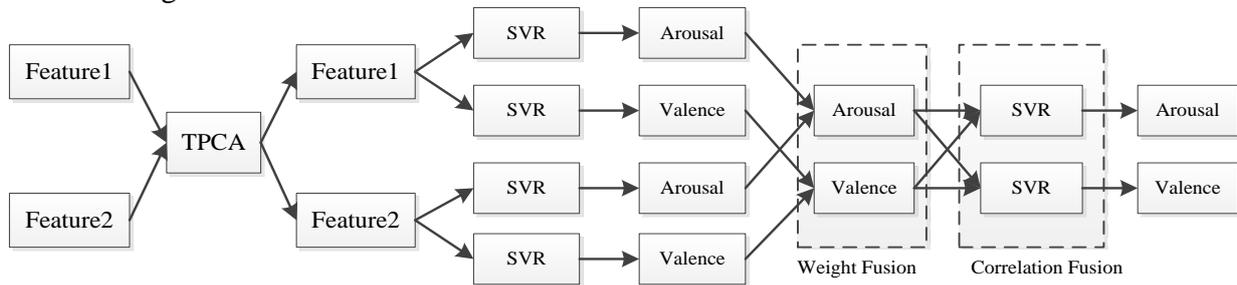


Fig. 1 Two-level fusion expression recognition framework based on TPCA

The proposed framework consists of the following steps. (1) Feature extraction. Gabor, LBP and LPQ features are adopted in this paper. (2) Feature dimensionality reduction. Two features are reduced dimensions by TPCA algorithm, respectively. (3) The first layer SVR training and prediction. SVR is used to train and predict the features after dimension reduction. (4) Weight fusion. Fusion of the same dimension predicted results of different features. (5) Correlation fusion. Firstly, two features are reduced dimensions by TPCA, respectively. Then, two features after dimension reduction of samples from training dataset are used for arousal and valence prediction by the several trained SVRs. The predicted arousal and valence values are put into two different SVR regressors for training after the weight fusion in the same dimension predicted results of different features. Finally, the final arousal and valence value can be predicted by the two SVRs.

4. Experiment

4.1 Datasets and Settings

In the following experiment, the RECOLA [13] and AVEC2013 [14] datasets are selected to evaluate the effectiveness of the proposed methods. The two datasets are cross-validated. Face detection is used by the face classifier from OpenCV. The image preprocessing is based on the clipping method in [15]. All the images are normalized to 120×120, then every image is converted to grayscale and performed histogram equalization. At last images are divided into 3×3 non-overlapping blocks after preprocessing. Gabor, LBP and LPQ features are extracted on each block. After feature extraction, TPCA dimensional reduction method is adopted performed on original features.

Three commonly used method of experimental evaluation are selected in this paper, which are Root Mean Square Error (RMSE), Pearson Correlation Coefficient (COR) and Mean Absolute Error (MAE). The parameter λ in the experiment is chosen empirically. The weighted weight of weighted fusion adopts the proportion of COR and adjusts appropriately according to the weight coefficient.

4.2 Experimental Results and Analysis

To evaluate the effective of the transfer learning algorithm TPCA, the traditional subspace PCA algorithm is compared. The experiments are performed on Gabor, LBP and LPQ respectively. The experimental results are shown in Fig 2, Fig 3.

Fig.2 and Fig.3 show that TPCA algorithm obtains better experimental results than the method of PCA, and the values of COR has significantly improved. The results show that when the two datasets do not satisfy the independent and identical distribution conditions, the performance of the PCA method is significantly degraded.

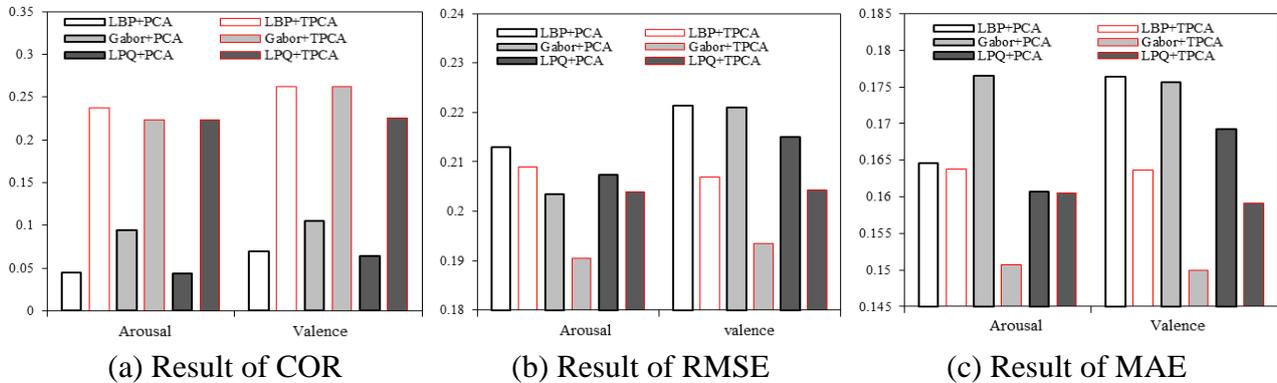


Fig. 2 Experiment results of AVEC2013 prediction RECOLA dataset on different features

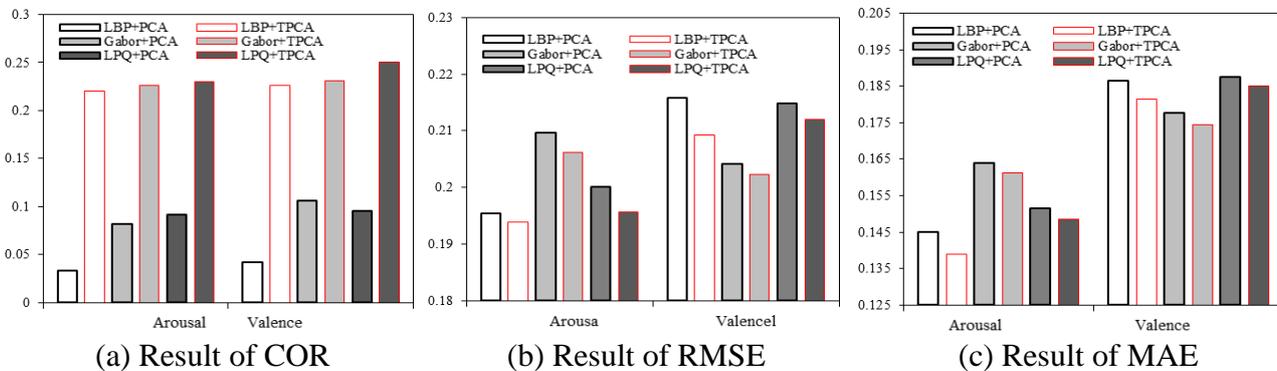


Fig. 3 Experiment results of RECOLA prediction AVEC2013 dataset on different features

Experiment of correlation fusion and without the correlation fusion between the A-V dimensions are compared. The experiments are performed on Gabor, LBP and LPQ. Experimental results are shown in Table 1, 2, 3 and 4.

The results of the experiment show that using correlation fusion method achieves better results for the same feature. What’s more, TPCA algorithm is better than the PCA algorithm, which further proves the effectiveness of the transfer learning algorithm. Compared with Fig.2, Fig.3, the results obtained from PCA and TPCA algorithm without correlation fusion are consistent with result of single SVR. This indicates that correlation fusion framework does not improve the experimental results without the A-V correlation fusion in single feature situation.

Table 1. Experimental results of AVEC2013 prediction RECOLA dataset based PCA method

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
LBP+PCA+no-correlation	0.20717	0.04430	0.16056	0.21412	0.07013	0.16820
LBP+PCA+correlation	<u>0.20673</u>	<u>0.04970</u>	<u>0.16024</u>	<u>0.21342</u>	<u>0.07924</u>	<u>0.16743</u>
Gabor+PCA+no-correlation	0.19453	0.09457	0.16901	0.20140	0.10572	0.15459
Gabor+PCA+correlation	<u>0.19137</u>	<u>0.10834</u>	<u>0.16614</u>	<u>0.19590</u>	<u>0.12436</u>	<u>0.14886</u>
LPQ+PCA+no-correlation	0.20153	0.04386	0.15697	0.20642	0.06405	0.15990
LPQ+PCA+correlation	<u>0.18460</u>	<u>0.06527</u>	<u>0.14986</u>	<u>0.19890</u>	<u>0.11484</u>	<u>0.15197</u>

Table 2. Experimental results of AVEC2013 prediction RECOLA dataset based TPCA method

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
LBP+TPCA+no-correlation	0.20148	0.23728	0.15748	0.20685	0.26308	0.16353
LBP+TPCA+correlation	<u>0.18862</u>	<u>0.25070</u>	<u>0.14904</u>	<u>0.20195</u>	<u>0.29416</u>	<u>0.15863</u>
Gabor+TPCA+no-correlation	0.18808	0.22378	0.14942	0.18917	0.26295	0.14345
Gabor+TPCA+correlation	<u>0.18796</u>	<u>0.25843</u>	<u>0.14895</u>	<u>0.18687</u>	<u>0.29140</u>	<u>0.14168</u>
LPQ+TPCA+no-correlation	0.19814	0.22313	0.15559	0.19832	0.22602	0.15389
LPQ+TPCA+correlation	<u>0.19414</u>	<u>0.23987</u>	<u>0.15292</u>	<u>0.19335</u>	<u>0.25463</u>	<u>0.14720</u>

Table 3. Experimental results of RECOLA prediction AVEC2013 dataset based PCA method

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
LBP+PCA+no-correlation	0.19192	0.03914	0.13949	0.21045	0.04225	0.18107
LBP+PCA+correlation	<u>0.18581</u>	<u>0.04852</u>	<u>0.12733</u>	<u>0.20784</u>	<u>0.05043</u>	<u>0.18001</u>
Gabor+PCA+no-correlation	0.20837	0.08146	0.16232	0.20266	0.10630	0.17537
Gabor+PCA+correlation	<u>0.20606</u>	<u>0.09556</u>	<u>0.15938</u>	<u>0.19728</u>	<u>0.11036</u>	<u>0.16981</u>
LPQ+PCA+no-correlation	0.19915	0.09122	0.15006	0.21180	0.09509	0.18430
LPQ+PCA+correlation	<u>0.19889</u>	<u>0.09599</u>	<u>0.14972</u>	<u>0.20933</u>	<u>0.10390</u>	<u>0.18163</u>

Table 4. Experimental results of RECOLA prediction AVEC2013 dataset based TPCA method

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
LBP+PCA+no-correlation	0.19137	0.22074	0.14225	0.21032	0.22661	0.18274
LBP+PCA+correlation	<u>0.18799</u>	<u>0.23643</u>	<u>0.13638</u>	<u>0.20916</u>	<u>0.24308</u>	<u>0.18147</u>
Gabor+PCA+no-correlation	0.20612	0.22648	0.16128	0.20236	0.23056	0.17443
Gabor+PCA+correlation	<u>0.19483</u>	<u>0.23445</u>	<u>0.14774</u>	<u>0.19884</u>	<u>0.23989</u>	<u>0.17029</u>
LPQ+PCA+no-correlation	0.19345	0.23024	0.14656	0.21134	0.25181	0.18460
LPQ+PCA+correlation	<u>0.19058</u>	<u>0.24235</u>	<u>0.14232</u>	<u>0.20865</u>	<u>0.26311</u>	<u>0.18179</u>

The two-level fusion method based on TPCA is performed to test feasibility of the framework in Fig 1, and also compare this method with one-feature method on the basis of A-V correlation fusion and TPCA. The experimental results are shown in Tables 5, 6, 7, 8, 9 and 10.

Table 5. The results of AVEC2013 prediction RECOLA dataset on Gabor and LBP features

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
Gabor+correlation-fusion	0.18796	0.25843	0.14895	0.18687	0.29140	0.14168
LBP+correlation-fusion	0.18862	0.25070	0.14904	0.20195	0.29416	0.15863
Gabor+LBP+two-level fusion	<u>0.18567</u>	<u>0.28798</u>	<u>0.14765</u>	<u>0.17862</u>	<u>0.32516</u>	<u>0.13211</u>

Table 6. The results of AVEC2013 prediction RECOLA dataset on Gabor and LPQ features

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
Gabor+correlation-fusion	0.18796	0.25843	0.14895	0.18687	0.29140	0.14168
LPQ+correlation-fusion	0.19414	0.23987	0.15292	0.19335	0.25463	0.14720
Gabor+LPQ+two-level fusion	0.17920	0.26752	0.14753	0.18000	0.32464	0.13556

Table 7. The results of AVEC2013 prediction RECOLA dataset on LBP and LPQ features

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
LBP+correlation-fusion	0.18862	0.25070	0.14904	0.20195	0.29416	0.15863
LPQ+correlation-fusion	0.19414	0.23987	0.15292	0.19335	0.25463	0.14720
LBP+LPQ+two-level fusion	<u>0.18304</u>	<u>0.25807</u>	<u>0.14770</u>	<u>0.18858</u>	<u>0.31070</u>	<u>0.14238</u>

Table 8. The results of RECOLA prediction AVEC2013 dataset on Gabor and LBP features

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
Gabor+correlation-fusion	0.19483	0.23445	0.14774	0.19884	0.23989	0.17029
LBP+correlation-fusion	0.18799	0.23643	0.13638	0.20916	0.24308	0.18147
Gabor+LBP+two-level fusion	<u>0.17718</u>	<u>0.28129</u>	<u>0.12419</u>	<u>0.19093</u>	<u>0.24947</u>	<u>0.16177</u>

Table 9. The results of RECOLA prediction AVEC2013 dataset on Gabor and LPQ features

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
Gabor+correlation-fusion	0.19483	0.23445	0.14774	0.19884	0.23989	0.17029
LPQ+correlation-fusion	0.19058	0.24235	0.14232	0.20865	0.26311	0.18179
Gabor+LPQ+two-level fusion	<u>0.17830</u>	<u>0.30519</u>	<u>0.12443</u>	<u>0.18848</u>	<u>0.27632</u>	<u>0.15957</u>

Table 10. The results of RECOLA prediction AVEC2013 dataset on LBP and LPQ features

Method	Arousal			Valence		
	RMSE	COR	MAE	RMSE	COR	MAE
LBP+correlation-fusion	0.18799	0.23643	0.13638	0.20916	0.24308	0.18147
LPQ+correlation-fusion	0.19058	0.24235	0.14232	0.20865	0.26311	0.18179
LBP+LPQ+two-level fusion	<u>0.17983</u>	<u>0.25100</u>	<u>0.12699</u>	<u>0.20437</u>	<u>0.26891</u>	<u>0.17673</u>

Those results show that two-level fusion method based on TPCA get the best experimental results. It can be concluded that the method proposed in this paper is superior to the single feature method.

5. Summary

There is less research on emotional dimension expression recognition in cross-dataset. The arousal and valence dimensional model is adopted in this paper. Firstly, the cross-dataset expression recognition problem is analyzed, then the positive correlation between arousal and valence is used to two-level fusion expression recognition method based on TPCA. In the experiment, two datasets are cross-validated. Experimental results have demonstrated the effectiveness and efficiency of transfer learning and the positive correlation between arousal and valence.

References

- [1] Mehrabian, A. (2008). Communication without words. *Communication Theory*, 193-200.
- [2] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2), 124.
- [3] Gunes, H. (2010). Automatic, dimensional and continuous emotion recognition.
- [4] Russel, J. A. (1980). Acircumplexmodelofafect. *Journal of Personalityand*.
- [5] Davitz, J. R. (1969). *The language of emotion*. Academic Press.
- [6] Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3), 273-294.

- [7] Yong Yang, Wenbo Huang, Yoo-sung Kim, et al. Facial expression recognition method based on MSVR and arousal-valence emotion model. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, accepted.
- [8] Lu, J., & Tan, Y. P. (2010). A doubly weighted approach for appearance-based subspace learning methods. *IEEE Transactions on Information Forensics and Security*, 5(1), 71-81.
- [9] He, X., Yan, S., Hu, Y., Niyogi, P., & Zhang, H. J. (2005). Face recognition using Laplacianfaces. *IEEE transactions on pattern analysis and machine intelligence*, 27(3), 328-340.
- [10] Lu, J., & Tan, Y. P. (2010). Regularized locality preserving projections and its extensions for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(3), 958-963.
- [11] Yan, H., Ang, M. H., & Poo, A. N. (2011, May). Cross-dataset facial expression recognition. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on* (pp. 5985-5990). IEEE.
- [12] Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500), 2323-2326.
- [13] Ringeval, F., Sonderegger, A., Sauer, J., & Lalanne, D. (2013). Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition* (pp.1-8).
- [14] Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., & Bilakhia, S., et al. (2013). AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. *ACM International Workshop on Audio/visual Emotion Challenge* (pp.3-10).
- [15] Shih, F. Y., & Chuang, C. F. (2004). Automatic extraction of head and face boundaries and facial features. *Information Sciences*, 158, 117-130.