

International Workshop on Internet of Smart Things (IST2021)
November 1-4, 2021, Leuven, Belgium

Low Rank Tucker Decomposition for 2D+3D Facial Expression Recognition

Yunfang Fu^{a,b,*}, Yujuan Deng^{a,b}, Yuekui Zhang^{a,b}, Zhengyan Yang^{a,b}, Ruili Qi^{a,b}

^a*School of Computer Science & Engineering, Shijiazhuang University, Shijiazhuang 050035, P.R. China*

^b*Hebei Province Internet of Things Intelligent Perception and Application Technology Innovation Center, Shijiazhuang 050035, China*

Abstract

In this paper, a novel low rank Tucker decomposition for 3D facial expression recognition (LRTDFER) approach is proposed, in which the core tensor is utilized to characterize the low-rank attribute among the samples combined with the factor matrices with the graph regularization embedding. At first, a model based on a 4D tensor is constructed from the facial expression data. By Tucker decomposing the constructed 4D tensor, a resulting core tensor and factor matrices in different tensor modes are utilized to characterize the low-rankness among samples. Because of the loss of information during modelling the 4D tensor, the missing data from partly observed facial expression data are recovered by embedding the tensor completion. Finally, the proposed model is handled and solved by adopting the alternating direction method of multipliers (ADMM). Meanwhile, the classification prediction of facial expressions are implemented by Multi-class-SVM. Numerical experiments are conducted on BU-3DFE database. The experiment results have been verified that our proposed approach is more competitive.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the Conference Program Chairs

Keywords: Low rank; Tucker decomposition; Facial expression recognition; Graph regularization embedding

1. Introduction

Expression recognition (FE), which is regarded as an important non-verbal way of human emotional communication and an important branch of automatic face analysis, has attracted more and more attention in the field of computer vision and pattern recognition, and plays an important role in today's scientific challenges.

In last decades, 2D facial expression recognition mainly focuses on 2D modal fusion of images and videos or multi-modal fusion of audio-visual data. Although the 2D FER methods based on face texture analysis have made great progress, they are largely affected by the changes of illumination and pose, which often appear in real scenes.

* Corresponding author.

E-mail address: fu_yunfang@126.com

With the rapid innovation of 3D data acquisition equipment, 3D data are insensitive to variation of illumination and head pose, which are exactly what robust 3D FER requires. Hence 3D FER has wide application prospects in related domains, such as human-computer interface (HCI), pattern recognition, computer vision, image retrieval, facial animation and psychological research [1]. With the BU-3DFE database [2] releasing by the public, researches on 3D Facial expression recognition (3D FER) have attracted enormous attentions [3]. The prototypical expressions include seven basic emotions, i.e. *Disgust (DI)*, *Sadness (SA)*, *Anger (AN)*, *Happiness (HA)*, *Surprise (SU)*, *Fear (FE)* and *Neutral (NE)*, which are widely recognized and keep consistent in different races and cultures.

The approaches based on the model [5] and the feature are regarded as two main streams [4, 5] categorized roughly by existing approaches in 3D FER. Approaches of the former make use of a generic 3D face template firstly, then a number of neutral samples are averaged from this template as the training, and is fitted to match a 3D face as the testing. Finally, the obtained coefficients or parameters from shape deformation in the course of the fitting are finally utilized for subsequent classifier prediction. While the latter often focus on sensitive geometry features of the facial expression from input 3D face scans. The feature-based approaches [3, 4] mainly contain the geometric feature-based, the local patch-based and 2D image mapping-based methods, such as Depth, Texture Map, Gaussian curvature, Shape index features, Mean curvature, Normal Map (NOM), and so on.

The approaches based on the model are attractive theoretically. Due to insufficient recognition features and changes in face mesh topology, they are prone to fatal errors in convergence, e.g. incurred by mouth opening. In addition, they often require high computation. The methods based on the feature are straightforward and often need low computations, while the discriminative power of utilized facial features are depended on substantially, and automatic or manual landmarks are mostly required. In recent years, deep learning based methods [6, 7] and the methods based on tensor decomposition [8, 9] are used and applied into 3D facial expression, and are obtained satisfied results.

However, among the methods above, it is not considered that the low-rank attribute among samples can be characterized by both the factor matrices with the graph regularization embedding and the core tensor. Hence, we propose a low rank Tucker decomposition approach which is applied into 2D+3D facial expression recognition (LRTDFER) in this paper, in which the core tensor is used to characterize the low-rank attribute among the samples combined with the factor matrices with the graph regularization embedding. At the first, a 4D tensor model is constructed from the facial expression data. After Tucker decomposing [10] the constructed 4D tensor, a resulting core tensor and factor matrices in different tensor modes are utilized to characterize the low-rankness among samples. Because of the loss of information in the course of modelling the 4D tensor, the missing data from partly observed facial expression data are recovered by embedding the tensor completion. Finally, the proposed model is handled and solved by using the alternating direction method of multipliers (ADMM). Meanwhile, the classification prediction of facial expressions are implemented by Multi-class-SVM. Numerical experiments are carried out on person-independent and gender-independent cases simultaneously on BU-3DFE database that is the most popularly employed for evaluating 3D FER approaches.

2. Preliminaries on Tensors

In this paper, lowercase letters (e.g., y), capital letters (e.g., Y) and calligraphic letters (e.g., \mathcal{Y}) represent vectors, matrices and tensors, respectively. Symbols $*$, \circ and \otimes are represented Hadamard product, outer, and the Kronecker, respectively.

2.1. Tucker Decomposition

Tucker decomposition is an effective method to extract the internal structure of multidimensional data. Given an N th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with its elements are represented by X_{i_1, i_2, \dots, i_N} ($1 \leq i_n \leq I_n$), then there always exist integers $R_1, \dots, R_n, \dots, R_N$ ($R_n \leq I_n, n = 1, 2, \dots, N$), a core tensor $\mathcal{G} = (\mathcal{G}_{r_1, \dots, r_N}) \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$, and matrices $U_1 \in \mathbb{R}^{I_1 \times R_1}, \dots, U_N \in \mathbb{R}^{I_N \times R_N}$, so that $\mathcal{X} = \mathcal{G} \times_1 U_1 \times_2 U_2 \cdots \times_N U_N$ when $U_n^T U_n = I_n$ ($n = 1, 2, \dots, N$) is satisfied. The Frobenius norm of \mathcal{X} is defined by $\|\mathcal{X}\|_F^2 = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} X_{i_1, i_2, \dots, i_N}^2$.

In tensor operations, $X_{(n)} \in \mathbb{R}^{I_n \times (I_1 \cdots I_{n-1} I_{n+1} \cdots I_N)}$ represents the mode- n unfolding of an N th-order tensor \mathcal{X} , where the mode- n folding is the procedure from $X_{(n)}$ to \mathcal{X} . The mode- n product of \mathcal{X} with a matrix U_n is denoted by $\mathcal{Y} = \mathcal{X} \times_n U_n$, where $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times R_n \times I_{n+1} \times \dots \times I_N}$. Meanwhile the mode- n unfolding of $\mathcal{Y} = \mathcal{X} \times_n U_n$ can be represented as $Y_{(n)} = U_n X_{(n)}$.

2.2. Graph regularization Extension

In order to improve the prediction accuracy of multi-dimensional tensor decomposition, the graph regularization framework [11] for tensor factorization is utilized to incorporate sample similarity into matrix factorization, in which the auxiliary information are exploited and regarded as similarity matrices as follows:

$$\min_{\mathcal{G}, \{U_n\}, \mathcal{X}, U_n^T U_n = I_n} \alpha \sum_{n=1}^N \text{Tr}(U_n^T L_n U_n) \quad \text{s.t.} \quad \mathcal{X} - \mathcal{G} \prod_{n=1}^N \times_n U_n = 0, \mathcal{X}(\Omega) = \mathcal{X}_0(\Omega) \quad (1)$$

where $\alpha \geq 0$, $\text{Tr}(\cdot)$ and $L_n = D_n - W_n$ represent a regularization constant, the matrix trace and the graph Laplacian matrix, respectively. Meanwhile W_n and D_n denote the weight matrix of the samples and the diagonal matrix, i.e., $(D_n)_{ii} = \sum_j (W_n)_{ij}$.

3. The Low Rank Tensor Approach

Now, 3D face scans and 2D face images from a samples are used, in which M kinds of features are extracted with its size of $I_1 \times I_2$ and are constructed into a 4D tensor \mathcal{X}_0 with its size of $\mathcal{X}_0 \in \mathbb{R}^{I_1 \times I_2 \times M \times N}$. Because of the information missed in the process of modeling the 4D tensor, an appropriate tensor \mathcal{X} will be utilized by recovering \mathcal{X}_0 . Meanwhile the orthogonal Tucker decomposition $\mathcal{X} = \mathcal{G} \prod_{n=1}^4 \times_n U_n$ is used for finding the factor matrices to project the 4D tensor for expression classification prediction.

It is well known that a Tucker decomposition of a high-order tensor has high computational cost, resulting in the ability of solving large-scale problems limited. Motivated by the challenge, we proposed a low rank Tucker decomposition approach for 2D+3D FER by replacing $\|\mathcal{X}\|_*$ with $\|\mathcal{G}\|_*$.

3.1. The Tensor Optimization Model

Theorem 1: Given two tensors: $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with n -rank $= (r_1, r_2, \dots, r_N)$, $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$ with $R_n \geq r_n$ ($n = 1, 2, \dots, N$). If $\mathcal{X} = \llbracket \mathcal{G}; U_1, \dots, U_N \rrbracket$ and $U_n^T U_n = I_{R_n}$ are satisfied, then $\|\mathcal{X}\|_* = \|\mathcal{G}\|_*$ where $\|\mathcal{G}\|_*$ and $\|\mathcal{X}\|_*$ are the trace norm of the core tensor \mathcal{G} and the tensor \mathcal{X} , respectively.

According to Theorem 1, the proposed approach is given as the general tensor optimization model in the following:

$$\min_{\mathcal{G}, \{U_n\}, \{G_{(n)}\}} \sum_{n=1}^4 \alpha_n \|\mathcal{G}_{(n)}\|_* + \gamma \sum_{n=1}^4 \text{Tr}(U_n^T L_n U_n) \quad \text{s.t.} \quad \mathcal{X} - \mathcal{G} \prod_{n=1}^4 \times_n U_n = 0, \mathcal{X}(\Omega) = \mathcal{X}_0(\Omega), U_n^T U_n = I, \quad (2)$$

where γ and α_n s denote a positive tradeoff parameter to compromise the effect of the two terms and the weight of $G_{(n)}$, respectively. The values of the term $\sum_{n=1}^4 \text{Tr}(U_n^T L_n U_n)$ are expected to be small so that discriminatively low dimensional subspace features can be better obtained from the constructed 4D tensor. \mathcal{X} is a recovered and reconstructed tensor. $G_{(n)}$ is the unfolding of the n -mode of \mathcal{G} . In order to distinguish \mathcal{G} and $G_{(n)}$, four auxiliary matrices $M_{(n)}$ s are introduced into (2) by replacing $G_{(n)}$ s, thus (2) is rewritten as:

$$\min_{\mathcal{G}, \{U_n\}, \{M_n\}} \sum_{n=1}^4 \alpha_n \|M_n\|_* + \gamma \sum_{n=1}^4 \text{Tr}(U_n^T L_n U_n) \quad \text{s.t.} \quad \mathcal{X} - \mathcal{G} \prod_{n=1}^4 \times_n U_n = 0, G_{(n)} = M_n, \mathcal{X}(\Omega) = \mathcal{X}_0(\Omega), U_n^T U_n = I, \quad (3)$$

3.2. Solution of the Optimization Problem

Obviously, problem (3) is nonlinear and non-convex including lots of constraints and variables. To solve this problem and ease the computation, the alternating direction method of multipliers (ADMM) [12] is utilized. Now, the augmented Lagrange function of (3) is defined as below by being given $(\mathcal{G}, \{U_n\}, \mathcal{X}, \{M_n\}, \mathcal{P}_x, \{P_n\}, \mathcal{P}_t, \mu)$:

$$\begin{aligned} \mathcal{L}_\mu(\mathcal{G}, \{U_n\}, \{M_n\}, \mathcal{P}_x, \{P_n\}, \mathcal{P}_t) = & \sum_{n=1}^4 \alpha_n \|M_n\|_* + \gamma \sum_{n=1}^4 \text{Tr}(U_n^T L_n U_n) + \langle \mathcal{X} - \mathcal{G} \prod_{n=1}^4 \times_n U_n, \mathcal{P}_x \rangle + \frac{\mu}{2} \|\mathcal{X} - \mathcal{G} \prod_{n=1}^4 \times_n U_n\|_F^2 \\ & + \sum_{n=1}^4 \langle G_{(n)} - M_n, P_n \rangle + \frac{\mu}{2} \sum_{n=1}^4 \|G_{(n)} - M_n\|_F^2 + \langle \mathcal{X}(\Omega) - \mathcal{X}_0(\Omega), \mathcal{P}_t \rangle + \frac{\mu}{2} \|\mathcal{X}(\Omega) - \mathcal{X}_0(\Omega)\|_F^2, \end{aligned} \quad (4)$$

where $\mu > 0$ is a scalar, U_n must satisfy $U_n^T U_n = I_n$ ($n = 1, 2, 3, 4$), \mathcal{P}_x , P_n s and \mathcal{P}_t are the Lagrange multipliers,.

3.2.1. Solution of \mathcal{G}

By fixing other parameters, \mathcal{G} can be solved by :

$$\min_{\mathcal{G}} \quad \frac{\mu}{2} \|\mathcal{G} - \mathcal{Q}\|_F^2 + \frac{\mu}{2} \sum_{n=1}^4 \|G_{(n)} - M_n + \frac{P_n}{\mu}\|_F^2. \quad (5)$$

where $\mathcal{Q} = (\mathcal{X} + \frac{\mathcal{P}_x}{\mu}) \prod_{n=1}^4 \times_n U_n^T$, a closed-form solution can be given:

$$\mathcal{G} = \frac{\mu(\mathcal{Q} + \mathcal{N})}{\mu(N + 1)} \quad (6)$$

where $\mathcal{N} = \sum_{n=1}^4 \text{fold}(M_n - P_n)$ and $N = 4$.

3.2.2. Solution of U_n 's

By fixing U_k ($k \neq n$) and other parameters, U_n ($1 \leq n \leq 4$) can be updated by

$$\min_{U_n^T U_n = I} \quad \frac{\mu}{2} \|\mathcal{X} - \mathcal{G} \prod_{n=1}^4 \times_n U_n + \frac{\mathcal{P}_x}{\mu}\|_F^2 + \gamma \text{Tr}(U_n^T L_n U_n). \quad (7)$$

Utilizing the equation $\mathcal{Y} = \mathcal{S} \times_n V \Leftrightarrow Y_{(n)} = V S_{(n)}$, (7) turns to the problem as below:

$$\max_{U_n^T U_n = I} \quad \langle U_n, A_n \rangle. \quad (8)$$

where $A_n = \mu(\mathcal{X} + \frac{\mathcal{P}_x}{\mu})_{(n)} \mathcal{G}_{(-n)}^T - \gamma L_n U_n$, and $\mathcal{G}_{(-n)} = \mathcal{G} \prod_{k=1, k \neq n}^4 \times_k U_k^T$. (8) can be solved easily by using the von Neumann's trace inequality [13] by

$$U_n = B_n C_n, \quad (9)$$

where $A_n = B_n D_n C_n$ is the SVD decomposition of A_n .

3.2.3. Solution of M_n 's

By fixing M_k ($k \neq n$) and other parameters, M_n ($1 \leq n \leq 4$) can be updated by

$$\min_{M_n} \quad \alpha_n \|M_n\|_* + \frac{\mu}{2} \|G_{(n)} - M_n + \frac{P_n}{\mu}\|_F^2. \quad (10)$$

Hence, a closed-form solution is achieved in (10) as follows:

$$M_n = \Theta_{\frac{\alpha_n}{\mu}}(G_{(n)} + \frac{P_n}{\mu}) \quad (11)$$

where $\Theta_\delta(W) = D S_\delta(\Sigma) V^T$, in which the singular value decomposition (SVD) of any matrix W denotes $W = D \Sigma V^T$, and $S_\delta(\Sigma_{ij}) = \max(0, \Sigma_{ij} - \delta)$ represents the soft-thresholding operator.

3.2.4. Solution of \mathcal{X}

By fixing other parameters, \mathcal{X} is achieved easily to implement with the following method:

$$\begin{cases} \bar{\Omega}(\mathcal{X}) = \bar{\Omega}(\mathcal{G} \prod_{n=1}^4 \times_n U_n - \frac{\mathcal{P}_x}{\mu}), \\ \Omega(\mathcal{X}) = \Omega(\mathcal{G} \prod_{n=1}^4 \times_n U_n - \frac{\mathcal{P}_x}{\mu} + \mathcal{X}_0 - \frac{\mathcal{P}_t}{\mu}). \end{cases} \quad (12)$$

The proposed algorithm can be summarized combined with the ADMM framework in Algorithm 1.

Algorithm 1 Solving Problem (3) by ADMM

Input: A tensor $\mathcal{X}_0 \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_4}$; Parameters $\alpha_n, \gamma, t_{\max}$;

Output: Factor matrices $\{U_n\}_{n=1}^4$;

Step 0 Initialization: choose $\{U_n^0\}_{n=1}^4, \mathcal{G}^0, \mathcal{X}^0, \rho, \mu^0, \{M_n^0\}_{n=1}^4$ and set $t = 0$;

Step 1 Update \mathcal{G}^t by (6);

Step 2 Update $\{U_n^t\}_{n=1}^4$ by (9);

Step 3 Update \mathcal{X}^t by (12);

Step 4 Update M_n^t by (11);

Step 5 Update multipliers by $\mathcal{P}_x^t = \mathcal{P}_x^{t-1} + \mu^{t-1}(\mathcal{X}^t - \mathcal{G}^t \prod_{n=1}^4 \times_n U_n^t)$, $P_n^t = P_n^{t-1} + \mu^{t-1}(G_{(n)}^t - M_n^t)$ and $\mathcal{P}_t^t = \mathcal{P}_t^{t-1} + \mu^{t-1}(\mathcal{X}^t(\Omega) - \mathcal{X}_0(\Omega))$;

Step 6 Let $\mu^t = \rho \mu^{t-1}$;

Step 7 $t=t+1$; While some stop criteria are not satisfied, go to Step 1.

In Algorithm 1, the stop criteria can be set as belows: $\|\Omega(\mathcal{X}^t - \mathcal{G}^{t-1} \prod_{n=1}^4 \times_n U_n^{t-1})\|_F / \|\mathcal{X}_0\|_F < \zeta$ with some sufficiently small $\zeta > 0$, or the iteration process will be terminated when $t > t_{\max}$.

4. Experiment Results

Experiments are implemented in the BU-3DFE database. 90% of 100 subjects are selected randomly regarded as the training, while the rest as the testing, and we repeat the experiments for 10 times and treat Multi-Class-SVM as the classifier in all experiments.

4.1. Experimental Details

Two parts are discussed in this subsection: one is the nine features extracted, the other is the parameters in Algorithm 1. The nine features extracted can be seen in [8]. $\alpha_n (n = 1, 2, 3, 4)$ are set to be 0.1, 0.1, 10, 0.01, respectively, and γ is set to be 100, and ζ is set to be 1e-4 in the experiments.

4.2. Experimental Evaluation

The average confusion matrix of recognition accuracies of six expressions are given in Table 1. It is found obviously that happiness and surprise expressions can be easily recognized because of their high facial muscle deformation, while the recognition of disgust and fear are obtained are more difficult.

Table 1. Confusion Matrix for six facial expressions on BU-3DFE for 10 Times.

%	AN	DI	FE	HA	SA	SU
AN	96.42	1.15	0.65	0.00	1.78	0.00
DI	2.71	92.13	3.53	0.42	0.00	1.21
FE	0.32	1.63	89.35	6.36	1.50	0.84
HA	0.00	0.05	1.04	98.91	0.00	0.00
SA	2.46	0.91	1.87	0.00	94.76	0.00
SU	0.00	0.49	0.21	0.42	0.00	99.30
Accuracy(%)	95.15					

Table 2. The comparisons with the state-of-the-art on BU-3DFE database.

Method	Accuracy
Soyel et al. [14]	95.10
Wang et al. [15]	83.60
Zhao et al. [16]	82.30
Fu et al. [4]	85.802
Ours	95.15

At the same time, some state-of-the-art methods are compared with our proposed approach, the comparisons are given in Table 2, and we can find that our proposed approach are more competitive from Table 2.

5. Conclusions

In this paper, we have proposed a low rank Tucker decomposition model and applied into 2D+3D facial expression recognition (IRTDFER). At first, a 4D tensor is built by extracting nine kinds of features from 2D face images and 3D face scans. Secondly, low rank attribute of 4D tensor among samples is characterized by a core tensor and four factor matrices by Tucker decomposing the 4D tensor. Thirdly, a tensor completion framework is embedded because of the data missed in the course of the 4D tensor modelling. Finally, the ADMM algorithm is utilized to solve the proposed optimization model. The numerical experiments have been verified that the proposed approach is more competitive. In the future, more effective low rank algorithms based on Tucker decomposition are required, which is one topic of our future research.

References

- [1] T. Fang, X. Zhao, O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, "3D/4D facial expression analysis: An advanced annotated face model approach," *Image Vis. Comput.*, vol. 30, no. 10, pp. 738–749, 2012.
- [2] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Int. Conf. Automat. Face Gesture Recog.*, 2006, pp. 211–216.
- [3] C. A. Corneanu, M. Oliu, J. F. Cohn, and S. Escalera, "Survey on rgb, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1548–1568, 2016.
- [4] Y. Fu, Q. Ruan, G. An, and Y. Jin, "Fast nonnegative tensor factorization based on graph-preserving for 3D facial expression recognition," in *IEEE Int. Conf. Signal Process.*, 2017, pp. 292–297.
- [5] Q. Zhen, D. Huang, Y. Wang, and L. Chen, "Muscular movement model-based automatic 3D/4D facial expression recognition," *IEEE Trans. Multimedia*, vol. 18, no. 7, pp. 1438–1450, 2016.
- [6] Z. Chen, H. Di, Y. Wang, and L. Chen, "Fast and light manifold cnn based 3d facial expression recognition across pose variations," in *2018 ACM Multimedia Conference*, 2018, pp. 229–238.
- [7] I. H. Trimech, A. Maalej, and N. Amara, "Data augmentation using non-rigid cpd registration for 3d facial expression recognition," in *2019 16th International Multi-Conference on Systems, Signals Devices (SSD)*, 2019, pp. 164–169.
- [8] Y. Fu, Q. Ruan, Z. Luo, Y. Jin, G. An, and J. Wan, "Ferlrtc: 2d+3d facial expression recognition via low-rank tensor completion," *Signal Processing*, vol. 161, pp. 74–88, 2019.
- [9] Y. Fu, Q. Ruan, and Y. Jiang, "Sparse and low-rank tucker decomposition with its application to 2D+3D facial expression recognition," in *2020 15th IEEE International Conference on Signal Processing (ICSP)*, 2020, pp. 37–42.
- [10] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [11] Y. Liu, F. Shang, W. Fan, J. Cheng, and H. Cheng, "Generalized higher order orthogonal iteration for tensor learning and decomposition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. PP, no. 99, pp. 1–13, 2015.
- [12] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [13] L. Mirsky, "A trace inequality of john von neumann," *Monatshefte Für Mathematik*, vol. 79, no. 4, pp. 303–306, 1975.
- [14] H. Soyel and H. Demirel, "Facial expression recognition using 3D facial feature distances," in *Int. Conf. Image Anal. Recog.*, 2007, pp. 831–838.
- [15] J. Wang, L. Yin, X. Wei, and Y. Sun, "3D facial expression recognition based on primitive surface feature distribution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 2, 2006, pp. 1399–1406.
- [16] X. Zhao, D. Huang, E. Dellandra, and L. Chen, "Automatic 3D facial expression recognition based on a bayesian belief net and a statistical facial feature model," in *Int. Conf. Pattern Recog.*, 2010, pp. 3724–3727.