Emotion recognition from facial images with arbitrary views

Xiaohua Huang
huang.xiaohua@ee.oulu.fi
Guoying Zhao
gyzhao@ee.oulu.fi
Matti Pietikäinen
mkp@ee.oulu.fi

Center for Machine Vision Research
Department of Computer Science and Engineering
University of Oulu
Oulu, Finland

Facial expression recognition has been predominantly utilized to analyze the emotional status of human beings. In practice nearly frontal-view facial images may not be available. Therefore, a desirable property of facial expression recognition would allow the user to have any head pose. We show in our paper a new method to recognize arbitrary-view facial expressions by using discriminative neighborhood preserving embedding and multi-view concepts.

Given n training images with C classes, denoted as $\mathbf{X} = [\vec{x}_1, \ldots, \vec{x}_n] \in R^{\mathrm{D}\times \mathrm{n}}$, based on class labels, we construct the intra-class and inter-class sets Ω_p^{wi} and Ω_p^{bw} of $\vec{x}_p$, where $p = 1, \ldots, \mathrm{n}$. The former contains intra-class $\vec{x}_q$ ($q = 1, \ldots, \mathrm{k}^{\mathrm{wi}}$) nearest neighboring to $\vec{x}_p$, while the latter contains inter-class $\vec{x}_r$ ($\mathrm{r} = 1, \ldots, \mathrm{k}^{\mathrm{bw}}$) nearest neighboring to $\vec{x}_p$. We further design an intrinsic graph $\mathbf{G}^{\mathrm{wi}} = \{\mathbf{X}, \mathbf{V}^{\mathrm{wi}}\}$ that preserves the intrinsic structure of intra-class samples, and a penalty graph $\mathbf{G}^{\mathrm{bw}} = \{\mathbf{X}, \mathbf{V}^{\mathrm{bw}}\}$ that describes the margin across inter-class boundaries.

(1) To preserve the similarity of intra-class samples, the similarity matrix $\mathbf{V}^{\mathrm{wi}}$ for the intrinsic graph $\mathbf{G}^{\mathrm{wi}}$ can be obtained by minimizing the following formulation,

$$\varepsilon^{\mathrm{wi}}(\mathbf{X}) = \sum_p \| \vec{x}_p - \sum_{\vec{x}_q \in \Omega_p^{\mathrm{wi}}} v_{p,q}^{\mathrm{wi}} \vec{x}_q \|^2, \tag{1}$$

with the constraint $\sum_q v_{p,q}^{\mathrm{wi}} = 1$, where $v_{p,q}^{\mathrm{wi}}$ in $\mathbf{V}^{\mathrm{wi}}$ is the weight of the edge from $\vec{x}_p$ to $\vec{x}_q$.

Given $\mathbf{U} \in R^{\mathrm{D}\times \mathrm{d}}$, the sample $\vec{x}_p$ is transformed to this space via $\vec{y}_p = \mathbf{U}^T \vec{x}_p \in R^{\mathrm{d}\times 1}$, where d is the dimensionality of this space. Therefore, the sample $\vec{y}_p$ can be represented as a linear combination of its neighbors with the corresponding coefficients v_{pq}^{wi} as follows,

$$\varepsilon^{\mathrm{wi}}(\mathbf{Y}) = \sum_p \| \vec{y}_p - \sum_{\vec{y}_q \in \Omega_p^{\mathrm{wi}}} v_{p,q}^{\mathrm{wi}} \vec{y}_q \|^2 = \mathbf{U}^T \mathbf{S}^{\mathrm{wi}} \mathbf{U}, \tag{2}$$

where $\mathbf{S}^{\mathrm{wi}} = \mathbf{X}(\mathbf{I} - \mathbf{V}^{\mathrm{wi}})(\mathbf{I} - \mathbf{V}^{\mathrm{wi}})^T \mathbf{X}^T$ represents the local geometric structure of intra-class samples.

(2) The penalty similarity matrix $\mathbf{V}^{\mathrm{bw}}$ can be obtained by minimizing the following formulation,

$$\varepsilon^{bw}(\mathbf{X}) = \sum_p \| \vec{x}_p - \sum_{\vec{x}_r \in \Omega_p^{\mathrm{bw}}} v_{p,r}^{\mathrm{bw}} \vec{x}_r \|^2, \tag{3}$$

with the constraint $\sum_r v_{p,r}^{\mathrm{bw}} = 1$, where $v_{p,r}^{\mathrm{bw}}$ in $\mathbf{V}^{\mathrm{bw}}$ represents the weight of the edge from $\vec{x}_p$ to $\vec{x}_r$ with different class labels.

To maximize the boundary of samples of different class labels, $\mathbf{U}$ can be obtained by maximizing the following function,

$$\varepsilon^{bw}(\mathbf{Y}) = \sum_p \| \vec{y}_p - \sum_{\vec{y}_r \in \Omega_p^{\mathrm{bw}}} v_{p,r}^{\mathrm{bw}} \vec{y}_r \|^2 = \mathbf{U}^T \mathbf{S}^{\mathrm{bw}} \mathbf{U}, \tag{4}$$

where $\mathbf{S}^{\mathrm{bw}} = \mathbf{X}(\mathbf{I} - \mathbf{V}^{\mathrm{bw}})(\mathbf{I} - \mathbf{V}^{\mathrm{bw}})^T \mathbf{X}^T$ represents the local geometric structure of inter-class samples.

Maximum margin criterion [3] was proposed to maximize the margin between classes after dimensionality reduction. Additionally, it does not suffer from the small sample size problem. $\mathbf{U}$ can be obtained accordingly by the following function,

$$\mathbf{U} = \arg\max \{\mathbf{U}^T \mathbf{S}^{\mathrm{bw}} \mathbf{U} - \eta \mathbf{U}^T \mathbf{S}^{\mathrm{wi}} \mathbf{U}\}, \tag{5}$$

with the constraint $\mathbf{U}^T \mathbf{X} \mathbf{X}^T \mathbf{U} = 1$, where η is the balancing factor that adjusts the second term to ensure a positive objective function.

In practice, multi-view facial expression recognition aims to recognize facial expressions with arbitrary views. Here, we simply suppose that there are N views for each sample. Given the samples with the ith view ($i \in \{1, \ldots, \mathrm{N}\}$), they are denoted as $\mathbf{X}_i$. From Eqn.5, we can obtain the formula $\mathbf{A}_i = (\mathbf{S}_i^{\mathrm{bw}} - \eta_i \mathbf{S}_i^{\mathrm{wi}})$ of the ith view with respect to $\mathbf{X}_i$. Thus Eqn.5 can be revised as follows,

$$[\mathbf{U}_1, \ldots, \mathbf{U}_{\mathrm{N}}] = \arg\max \sum_{i=1}^{\mathrm{N}} \mu_i \mathbf{U}_i^T \mathbf{A}_i \mathbf{U}_i, \tag{6}$$

with constraints $\sum_i \mathbf{U}_i^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{U}_i = 1$, where the positive term μ_i is included to bring the balance between multiple objectives. Eqn.6 embeds the respective objective function of each view into one common function of all views, but still loses the correlation of samples with distinct views.

Multiset canonical correlation analysis (MCCA) [4] was proposed to search the correlation vector for multiple sets. According to MCCA and the formulation of closeness in [1], we define the objective function that maximizes the correlation between samples with the same expression label yet on all views and meanwhile minimizes the covariance of samples with the same view label, as follows,

$$[\mathbf{U}_1, \ldots, \mathbf{U}_{\mathrm{N}}] = \arg\max \sum_{i=1}^{\mathrm{N}} \sum_{j=1, j\neq i}^{\mathrm{N}} \mathbf{U}_i^T \mathbf{M}_i \mathbf{M}_j^T \mathbf{U}_j, \tag{7}$$

with the constraint $\sum_{i=1}^{\mathrm{N}} \mathbf{U}_i^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{U}_i = 1$, where $\mathbf{M}_i$ is the class mean matrix of samples on the ith view.

Based on two objective functions (6) and (7), we can obtain the completed formulation as follows,

$$[\mathbf{U}_1, \ldots, \mathbf{U}_{\mathrm{N}}] = \arg\max \sum_{i=1}^{\mathrm{N}} \{\mu_i \mathbf{U}_i^T \mathbf{A}_i \mathbf{U}_i + \sum_{j=1, j\neq i}^{\mathrm{N}} \alpha_{i,j} \mathbf{U}_i^T \mathbf{M}_i \mathbf{M}_j^T \mathbf{U}_j\}, \tag{8}$$

with the constraint $\sum_{i=1}^{\mathrm{N}} \beta_i \mathbf{U}_i^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{U}_i = 1$, where μ_i, $\alpha_{i,j}$ and β_i are balancing parameters in which adjust the importance of terms in the objective function and constraint item. Through this formula, we can obtain discriminative feature space of facial expression $\mathbf{U}_i$ in each view.

Emotion Classification: Our aim is to match two face images with the same or different facial expression label in different views. Partly motivated by cross-view classification [2], we design the mean-correlation maximization classifier to classify the sample $\vec{x}$ as follows,

$$h(\vec{x}) = \max_c (\mathrm{mean}_c (\max_i \{\mathrm{corr}(\mathbf{U}_i^T \mathbf{X}_{i,c}, \mathbf{U}_i^T \vec{x})\}_{i=1}^{\mathrm{N}})), \tag{9}$$

where $\mathbf{X}_{i,c}$ represents training samples of the cth facial expression label with the ith view, *corr* represents Pearson's linear correlation coefficient operator, N is the number of views, *mean* and *max* are the mean and maximum value operator, respectively.

Conclusion: Our method firstly captures the discriminative property of inter-class samples. In addition, it explores the closeness of intra-class samples with arbitrary view in a low-dimensional subspace.

[1] S. Abhishek, K. Abhishek, and D. Hal. Generalized multiview analysis: a discriminative latent space. In *Proc. CVPR*, pages 2160–2167, 2012.

[2] A. Li, S. Shan, and W. Gao. Coupled bias-variance trade off for cross pose face recognition. *IEEE Trans. Image Processing*, 21(1):305–315, 2012.

[3] H. Li, T. Jiang, and K. Zhang. Efficient and robust feature extraction by maximum margin criterion. In *Proc. NIPS*, pages 157–165, 2003.

[4] A. Nielsen. Multiset canonical correlation analysis and multispectral, truly multitemporal remote sensing data. *IEEE Trans. Image Processing*, 11(3):293–305, 2002.