Smile recognition based on the fusion of the face texture feature and the mouth shape feature

Li Yuanzheng

College of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an, Shanxi, China, 710054

Email: liyuanzheng@tom.com

Keywords: Smile recognition, Discriminative CCA, Feature fusion, LBP, HOG.

Abstract. Smile recognition has recently attracted significant attention due to the increasing accessibility in some digital products. Traditionally, the mouth feature has been extracted for smile recognition. However, it only extracts the features in the mouth region, ignoring the fact that face feature contribute to characterize expressions. In this paper, in order to overcome the drawbacks and make full use of the merits of both features, we propose a novel feature fusion approach for the smile recognition. In our approach, the mouth features are described by HOG descriptor and the face features are represented by Local Binary Pattern (LBP) histograms. Then, the two features are fused by discriminative canonical correlation analysis (DCCA). Robust Support Vector Machine (SVM) classifier is used for classification in our experiment and satisfactory results of experiments based on two databases are obtained finally.

1. Introduction

In recent years, with the rapid development of digital products and the growing demands from application, increasing attention has been paid to smile recognition from areas of lab research to actual application. For example, one application is to automatically detect smiles when taking photos, and this technique has been commercialized as a new function of digital cameras. As a result, improving the recognition performance of such digital systems has become an essential task.

Smile is regarded as an important expression, and for smile recognition, an essential task is to extract the features from face image efficiently, which is the principle problem of expression recognition. Over the last decade many methods have been proposed, which can be divided into two categories, image-based method and feature-based method [1]. The former is to extract features from face regions of an image. For example, using 2-D DCT [2], LBP [3] and the Gabor [4] in the extracting of the whole face features for expression recognition. The later is to extract features from components of a face, such as nose, mouth, eyes and eyebrows. For instance, in [5] only the eye shape changes are extracted.

However, it has also been noticed that both the image-based and the feature-based strategies have their own limitations. Since the feature-based method only takes the contribution of components of a face into account, ignoring the fact that the underlying face structure of particular face determinate the unique way for facial expression. Conversely, the image-based method takes the whole face into account but fail to lay stress on some regions with more clues.

To overcome the above limitations and preserve their merits of the two kinds of features, an effective way is to find a proper fusion of the applicable features. For this reason, we introduce a fusion approach for the two kinds of features based on the analysis of the correlation within them. More formally, we call the fusion method as discriminative canonical correlation analysis (DCCA) [6]. First, given that it is normally deemed that the mouth region for smile recognition plays a much more important role in comparison with the other regions on a face, we extract the features on the mouth region and HOG operator is introduced for representing the shape of mouth region. Second, block-weighted LBP operator is used to extract features on the whole region of human face because of its low computation cost and exceptional texture description ability. Finally, the two features are
fused by DCCA. The experimental results based on two databases illustrate the superiority of the proposed method.

While regarding classifier design, SVM [7] is adopted here since it has been successfully applied to expression recognition.

The remainder of the paper is organized as follows: In section 2, the representations of the face and mouth features are discussed. Section 3 presents the fusion method, and in section 4 we expatiate upon the experiment based on person-independent. Finally, conclusions are given in section 5.

2. Feature representation

The face feature is represented by LBP histograms while the mouth feature is represented by HOG descriptor.

LBP operator, introduced by Ojala [8] in 1996, is capable of describing texture. In this paper, we use the extended version of LBP operator chiefly, denoted by LBP$_{p,R}$. Let $R$ denote the radius of a circle, the centre of which is the central pixel and the number of the pixels is specified by $P$ in the circle. The formula is denoted as

$$\text{LBP}_{p,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p$$

Where $s(x)$ is defined as $s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$, $g_p, g_c$ are gray value of the p-th and center pixels.

HOG descriptor can effectively express the shape features and have been used in many different problems in computer vision, such as person detection and face detection [9]. The algorithm for extracting HOG begins by counting occurrences of edge orientations in localized portions of an image. The image is divided into small spatial regions, called cells, and for each cell a histogram of edge orientations is computed. The histogram channels are evenly spread over 0 to 180 degrees or 0 to 360 degrees, depending on whether the gradient is unsigned or not.

However, the histogram counts alone usually could not give us satisfactory results. For better invariance to illumination, shadowing, etc., the histogram counts should be contrast-normalized. This can be realized by accumulating a measure of local histogram “energy” over somewhat larger spatial regions (“blocks”) and using the results to normalize all the cells in the block. The combination of these histograms then represents the descriptor.

3. Feature fusion

Following the methods stated above, we obtain two kinds of features to represent smile expression. The fact that both features describe the same pattern inspires us that there should be correlation that can describe the relevant information contained in these two features. And we depend on the correlation to perform feature fusion. In the following, we present the feature fusion based on the DCCA in detail.

3.1 The basic idea of DCCA

CCA [10], which studies the correlation between two information channels, has been employed as a powerful tool for feature fusion in the realization of recognition system. However, one limitation of CCA is that it is unable to utilize supervised information. To this end, we introduce a new supervised learning method, namely discriminative CCA (DCCA), into the smile recognition system. DCCA aims to seek pairs of projection vector $W_x = (w_{x1}, w_{x2}, \ldots, w_{xd})$ and $W_y = (w_{y1}, w_{y2}, \ldots, w_{yd})$ for feature extraction such that the within-class correlation is maximized and the between-class correlation is minimized. More formally, DCCA can be formulated as the following optimization problem [11]:
\[ \rho_{\text{dcca}} = \frac{w_x^T \tilde{C}_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}} (\tilde{C}_{xy} = C_w - \eta C_b) \]  

(2)

Where \( C_w \) and \( C_b \) denote the matrices of the within-class and between-class correlations respectively, \( \eta (\eta > 0) \) is a tunable parameter that indicates the relative significance of the within-class correlation \( w_x^T C_{w} w_y \) versus the between-class correlation \( w_x^T C_{b} w_y \).

Given two sets of mean-normalized samples as \( X = [x_1^{(i)}, \ldots, x_n^{(i)}] \in \mathbb{R}^{p \times n} \), \( Y = [y_1^{(i)}, \ldots, y_n^{(i)}, \ldots, y_1^{(c)}, \ldots, y_n^{(c)}] \in \mathbb{R}^{q \times n} \). The purpose of DCCA is to compute \( w_x \) and \( w_y \) such that the within-class correlation is maximized and the between-class correlation is minimized. The within-class correlation matrix \( C_w \) can be defined as

\[ C_w = \sum_{i=1}^{c} \sum_{k=1}^{n} \sum_{l=1}^{n} x_k^{(i)} y_l^{(i)^T} = XDY^T \]  

(3)

In contrast, the between-class correlation matrix \( C_b \) can be defined as

\[ C_b = \sum_{i=1}^{c} \sum_{j=1}^{c} \sum_{k=1}^{n} \sum_{l=1}^{n} x_k^{(i)} y_l^{(j)^T} = -C_w \]  

(4)

So the difference between \( C_w \) and \( C_b \) is just one negative sign. Thus the numerator of (2) turns into \( (1 + \eta) w_x^T C_w w_y \) and this optimization problem is independent of the parameter \( \eta \). This in fact indicates that when DCCA maximizes the within-class correlation, at the same time, the between-class correlation is automatically minimized. Hence, the optimization problem of DCCA can be reformulated as follows:

\[ \begin{pmatrix} \theta \\ C_w^T \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \lambda \begin{pmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} \]  

(5)

3.2 The fusion algorithm

Once the basis vector pairs \( W_x = (w_{x1}, w_{x2}, \ldots, w_{xd}) \) and \( W_y = (w_{y1}, w_{y2}, \ldots, w_{yd}) \), which correspond to the first \( d \) largest values \( \lambda \), are obtained, for anyone pair of inside sample \( x \) and \( y \), we can perform feature fusion as follows:

\[ z = \left( \begin{array}{c} W_x^T \\ 0 \end{array} \right) \begin{pmatrix} x \\ y \end{pmatrix} \]  

(6)

Where \( z \) is the fused feature for the subsequent recognition, and \( d \) satisfies the constraints \( d \leq c \) and \( d \leq \min(p, q) \) due to \( \text{rank}(D) \leq c \).

Now, we can summarize the step of our feature fusion method based on DCCA as Algorithm 1:

Algorithm 1. Feature fusion algorithm

Input: the face feature described by LBP histograms and the mouth feature represented by the HOG descriptor

Output: the fused feature \( z \)

Step1: Setting the matrix \( X \) and \( Y \).

Step2: According to Step1, compute the within-class correlation matrix \( C_w \) with Eq.(3).

Step3: Solving the eigenvalue problem of Eq.(5) and attaining \( W_x \) and \( W_y \).

Step4: with Eq.(6) the fused feature \( z \) is our result.
4. Experiments

In this paper, robust SVM classifier is adopted for classification, and our experiment is to assess the performance of the proposed approach. We use the proposed fusion method to conduct the face texture and the mouth shape features. In order to illustrate the improvement of the proposed method, we also perform smile recognition with the face texture feature described by the block-weighted LBP histogram, the mouth shape feature represented by the HOG descriptor, the combined feature sets of the two kinds of features in series and the fused features based on CCA for comparison. And to sufficiently test the accuracy of our approach, we performed smile recognition on the JAFFE and CKAUFCFE databases[12]. Table 1 indicates the recognition accuracy on the JAFFE and the CKAUCFE, respectively. The best results have been highlighted in boldface. As can be seen on this table, the recognition accuracy in our approach is better than other approaches.

Table 2 Performance evaluation of person-independent experiments based on JAFFE and CKAUCFE databases.

|       | JAFFE | CKAUCFE |
|-------|-------|---------|
| LBP   | 90.47%| 90.66%  |
| HOG   | 90.47%| 87.94%  |
| LBP-HOG | 91.67%| 92.22%  |
| CCA   | 91.67%| 91.83%  |
| DCCA  | 95.24%| 93.39%  |

5. Conclusions

In this paper, we propose a method that combines the face texture feature and the mouth shape feature to perform smile recognition. We select the block-weighted LBP histograms to describe the face texture feature and the HOG descriptor to represent the mouth shape feature. The discriminative canonical correlation analysis (DCCA) is employed to fuse the two kinds of features. The experimental results demonstrate the effectiveness and superiority of our approach. In the following, we expect that it also has great potential for other facial expression recognition task, which merits our future study.

Acknowledgements

This work is supported by the PhD starting capital of Xi’an University of Science and Technology (Program No. 2014QDJ020). Besides, this work was supported in part by the Science and Technology Plan of Shaanxi Province, China (Program No. 2014K06–37; 2015GY023), the Science and Technology Plan Industrial Application Technology Research and Development Project of Xi’an (Program No.CXY1440 (4)), the Xi’an City Beilin District Application Technology Research and Development Project (Program No.GX1417), the PhD starting capital of Xi’an University of Science and Technology (Program No.2015QDJ015).All the persons involved in the research projects are thanked for their help.

References

[1] H. H. Tsai, Y. S. Lai, Y. C. Zhang, Using SVM to design facial expression recognition for shape and texture features, International Conference on Machine Learning and Cybernetics(2010) 2697-2704.
[2] Y. G. Xiao, L. Ma, K. Khorasani, A new facial expression recognition technique using 2-d dct and neural networks based design tree, in: IEEE International Joint Conference on Neural Networks(2006) 2421-2428.

[3] C. F. Shan, S. G. Gong, P. W. McOwan, Robust facial expression recognition using local binary patterns, in: IEEE International Conference on Image Processing(2005)II-370-3.

[4] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, J. Movellan, Dynamics of facial expression extracted automatically from video. Image and Vision Computing 24(6), 2006, 615-625.

[5] I. Saeed, A. Wang, R. Senaratne, S. Halganuge, Using the active appearance model to detect driver fatigue, in: Proceedings of the 3th International Conference on Information and Automation for Sustainability(2007)124-128.

[6] T. K. Sun, S. C. Chen, Z. Jin, J. Y. Yang, Kernelized discriminative canonical correlation analysis, in: IEEE International Conference on Wavelet Analysis and Pattern Recognition(2007)1283-1287.

[7] Information on http://www.csie.ntu.edu.tw/cjlin/libsvm

[8] T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognition 29(1), 1996, 51-59.

[9] N. D. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition(2005)886-893.

[10] H. Hotelling, Relations between two sets of variates, Biometrika 34(4)(1936)321-377.

[11] Q. Sun, S. Zeng, Y. Liu, P. A. Heng, D. S. Xia, A new method of feature fusion and its application in image recognition, Pattern Recognition 38(12)(2005)2437-2448.

[12] Information on http://www.kasrl.org/jaffe.html.