LETTER

Facial Expression Recognition via Sparse Representation

Ruicong ZHI[†,††a)], **Qiuqi RUAN**[†], ***and*** **Zhifei WANG**[†], ***Members***

SUMMARY A facial components based facial expression recognition algorithm with sparse representation classifier is proposed. Sparse representation classifier is based on sparse representation and computed by L1-norm minimization problem on facial components. The features of "important" training samples are selected to represent test sample. Furthermore, fuzzy integral is utilized to fuse individual classifiers for facial components. Experiments for frontal views and partially occluded facial images show that this method is efficient and robust to partial occlusion on facial images.

key words: *sparse representation, fuzzy integral, decision level fusion, facial expression recognition*

1. Introduction

Facial expression plays a key role in non-verbal face-to-face communication. It is a challenging task to recognize the facial expression from a static image. Since the intrinsic features of the facial expressions always hide in very high-dimensional space, it is necessary to find the meaningful low-dimensional structure by dimensionality reduction for facial representation. In general, feature extraction methods represent facial features in either holistic or local ways. Holistic representation mainly preserves the texture of a whole face image, and it has a large capacity of representing a new face image. However, holistic representation suffers from heavy training and high redundancy. Local representation adopts local facial regions for feature extraction and focuses on the subtle diversities on a face. It can be computed very fast and occupies little memory. However, it performs poorly on new test image not belonging to the training set. Some studies show that facial expression recognition prefers non-holistic representation [1], while good results are still obtained by using holistic approaches [2]. Hence, it motivates us to exploit both of their benefit to develop a hybrid representation.

For facial expression classification, the commonly used Nearest Neighbor Classifier (NNC) [3] and Nearest Subspace Classifier (NSC) [4] locally identify a test sample based on the smallest residual which is measured by the similarity between test sample and each training sample (NNC) or each facial expression class (NSC). Thus, they do not take consideration of the knowledge of training samples of other facial classes. Recently, a sparse representation classifier was proposed. Each test sample can be approximated as linear combination of training samples of the same facial expression class. If the facial expression classes are large enough, the coefficients are sparse. As sparse representation classifier harnesses the knowledge of all facial classes, it is better for classification than NNC and NSC [5].

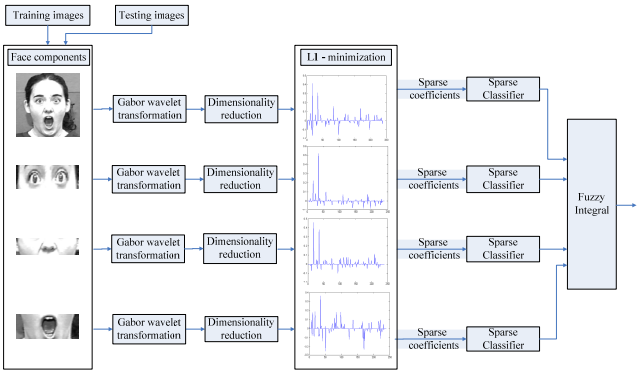


Fig. 1 Overall architecture of facial expression recognition using SRC.

Researches of human visual system show that human subjects respond to information around the eyes independently from variation around the mouth and they are able to recognize and distinguish isolated parts of faces. Therefore, we divide facial images into blocks and process each block independently. This paper is interesting from the following aspects: (1) facial expression features are extracted from both facial image and facial block images. It can efficiently combine the advantages of holistic features and local features; (2) sparse representation for facial classification takes consideration of all facial expression classes, so that to get global representation for test sample; (3) fusing both holistic and local facial features by fuzzy integral method. Our method is robust to partial occlusions on facial expression images. As partial occlusion can be treated as some special training samples, and they can be used to represent occluded testing samples. Furthermore, to eliminate the affect of different individuals to facial expression recognition, Gabor wavelet transformation is realized to extract more efficient facial features. The overall architecture of our method is shown in Fig. 1.

Manuscript received December 2, 2011.
Manuscript revised April 9, 2012.

[†]The authors are with Institute of Information Science, Beijing Jiaotong University, Beijing 100044, P. R. China.
[††]The author is with Food and Standardization, China National Institute of Standardization, Beijing 100088, P. R. China.
a) E-mail: zhirc@cnis.gov.cn

DOI: 10.1587/transinf.E95.D.2347

2. Sparse Representation Classifier

2.1 Sparse Representation via L1-Norm Minimization

Suppose there are n samples $\mathbf{A} = [a_1, a_2, \cdots, a_n]$ in a high-dimensional image space, each sample is denoted by a m-dimensional vector $a_i = [a_{i,1}, a_{i,2}, \cdots, a_{i,m}]^T$. There are C pattern classes, and the number of samples in kth class is n_k, i.e. $n = \sum_{k=1}^{C} n_k$. For a test sample y, it can be approximated by linear combination of selected training samples, that is

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = \mathbf{A}x \tag{1}$$

where $x = [x_1, x_2, \cdots, x_n]^T \in \mathbf{R}^n$ is coefficients vector whose entries are zero except some important training samples. Conventionally, the sparse solution of the underdetermined system is found with respect to the L0-norm which counts the number of non-zeros elements in coefficients vector. However, L0-norm minimization problem is NP-hard optimization problem, and it is hard to be approximated [6]. The L1-norm is often used as a penalty for sparsity as a proxy of L0-norm to avoid local minima in the optimization problem. The L1-norm minimization problem is defined as:

$$\hat{x}_1 = \arg\min \|x\|_1 \quad \text{subject to} \quad \mathbf{A}x = y \tag{2}$$

where $\|\cdot\|_1$ denotes the L1-norm, $\|x\|_1 = \sum_i |x_i|$. L1-norm minimization is typical convex optimization problem and can be solved by linear programming methods.

2.2 Sparse Representation Classifier

Sparse representation classifier (SRC) chooses the training samples which can best represent each test sample [5]. Let δ_k be the function that selects the sparse coefficients associated with the kth class. In $\delta_k(x)$, only the elements that are associated with the kth class are nonzero. Test sample y can be approximated by training samples from each class $\hat{y}_k = \mathbf{A}\delta_k(\hat{x}_1)$. Define the residual between test sample y and approximation $\hat{y}_k$ from each class samples as $r_k(y) = \|y - \mathbf{A}\delta_k(\hat{x}_1)\|_2$. Then test facial parts are assigned to the class that minimizes the residual between y and $\hat{y}_k$:

$$d(y) = \arg\min_k r_k(y) \tag{3}$$

2.3 Deal with Partial Occlusion

Until now, most of the facial expression experiments have been conducted in controlled laboratory conditions which do not always reflect the real-world condition. For example, human faces may be occluded by sunglasses, scarf, hands, etc. It is necessary to deal with partially occluded facial expression recognition problem. The existed holistic methods are usually not robust to occlusion. They treat the occluded part as normal facial images, and after feature extraction, the occluded part still affects the recognition results significantly.

Now we extend the sparse representation classifier to deal with partial occlusion. First, the face images are divided into three facial parts, namely eye part, nose part (including cheek) and mouth part. Together with the whole face, there are four blocks to be processed.

Suppose the size of facial blocks is $a \times b$. For training samples, the image matrix contains a set of facial parts matrice $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}, \mathbf{A}^{(4)} \in \mathbf{R}^{p \times n}$, where $p = a \times b$. Similarly, the test image can be represented by four facial blocks as $y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)} \in \mathbf{R}^p$. For each facial block, we find the sparse representation independently. The partially occluded blocks can be described by plenty of training samples blocks and a set of occluded blocks.

$$y^{(l)} = y_0^{(l)} + z_0^{(l)} = \mathbf{A}^{(l)} x_0^{(l)} + z_0^{(l)} \tag{4}$$

$$y^{(l)} = \begin{bmatrix} \mathbf{A}^{(l)} & \mathbf{I}^{(l)} \end{bmatrix} \begin{bmatrix} x_0^{(l)} \\ z_0^{(l)} \end{bmatrix} = \mathbf{B}^{(l)} w_0^{(l)} \tag{5}$$

where $\mathbf{B}^{(l)} = \begin{bmatrix} \mathbf{A}^{(l)} & \mathbf{I}^{(l)} \end{bmatrix} \in \mathbf{R}^{p \times (n+p)}$, $w_0^{(l)} = \begin{bmatrix} x_0^{(l)} & z_0^{(l)} \end{bmatrix}^T \in \mathbf{R}^{n+p}$. Similarly, the sparse solution can be obtained by the following L1-norm minimization problem:

$$\hat{w}_1^{(l)} = \arg\min \left\| w^{(l)} \right\|_1 \quad \text{subject to} \quad \mathbf{B}^{(l)} w^{(l)} = y^{(l)} \tag{6}$$

To classify the test sample, the residual is calculated as follows:

$$\begin{aligned} r_i^{(l)}(y) &= \left\| y^{(l)} - \mathbf{A}^{(l)} \delta_i \left(\hat{x}_1^{(l)} \right) \right\|_2 \\ &= \left\| y^{(l)} - \hat{z}_1^{(l)} - \mathbf{A}^{(l)} \delta_i \left(\hat{x}_1^{(l)} \right) \right\|_2 \end{aligned} \tag{7}$$

3. Fuzzy Integral

Fusing individual classifiers by fuzzy integral can be interpreted as simultaneously considering the classification results and the weights of the individual classifiers to reach a final classification result. Let $A = [a_1, a_2, \cdots, a_n]$ be a set of sources, $h_k(a_i)$ be the membership grade of a_i to the kth class, and g be the fuzzy measure. The fuzzy integral is defined as

$$\int_A h_k(a) \circ g(\cdot) = \sup_{\kappa \in [0,1]} [\min(\kappa, g(\{a \mid h_k(a) \geq \kappa\}))] \tag{8}$$

Choquet fuzzy integral [7] was used to fuse individual classifiers. If the values of $h_k(\cdot)$ are ordered in a non-degreasing order, $h_k(a_1) \leq h_k(a_2) \leq \cdots \leq h_k(a_n)$, then the Choquet fuzzy integral is defined as follows:

$$\int_A h_k(a) dg(\cdot) = \sum_{i=1}^{n} [h_k(a_i) - h_k(a_{i-1})] g(A_i) \tag{9}$$

where $A_i = [a_1, a_2, \cdots, a_i]$ denotes a subset of elements of the whole dataset. $g(A_i)$ can be calculated recursively in the form

$$\begin{aligned} g(A_1) &= g(\{a_1\}) = g^1 \\ g(A_i) &= g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}) \end{aligned} \tag{10}$$

where g^i is the value of the fuzzy density function, λ is the root of the equation $\lambda + 1 = \prod_{i=1}^{n} \left(1 + \lambda g^i\right)$.

4. Experimental Results

The proposed method is verified with application to facial expression recognition, including classifying frontal facial expression images and partially occluded facial expression images. The experiments are conducted on the commonly used Cohn-Kanade facial expression database [8]. As some subjects in CK database show less than six facial expressions, we use a subset of thirty subjects with six basic facial expressions (surprise, sadness, fear, disgust, anger, and happiness). For each expression of a subject, the last eight frames are selected as static images. The images are manually cropped to a central face image and resized to 120×120.

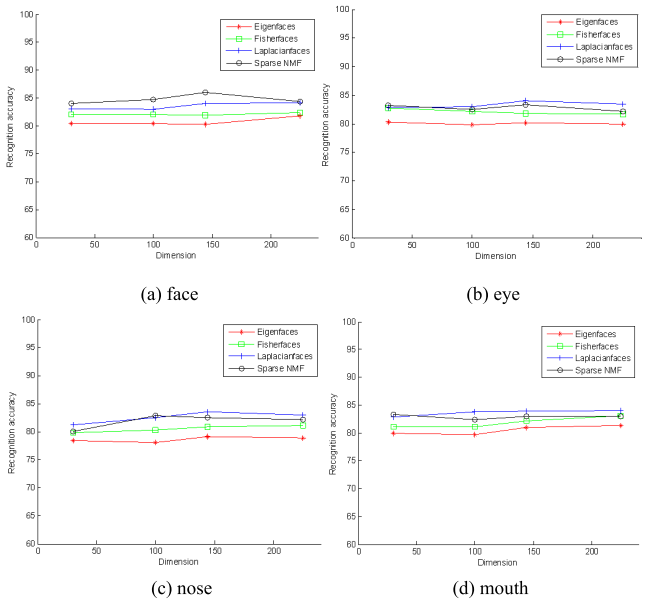
4.1 Gabor Wavelet Transformation

Gabor wavelet allows description of spatial frequency structure in the image, and it exhibits desirable characteristics of local spatiality and orientation selectivity [9]. As Gabor features are highly correlated and redundant among neighboring pixels, it is sufficient to extract Gabor features from some landmarks on facial images. For whole facial images, we select downsample points in eight rows and six columns with fixed distance, so that we get a 48×40 dimensional Gabor feature vector. Then the Gabor features are utilized as input of dimensionality reduction and classification block.

4.2 Facial Expression Recognition without Occlusion

Facial expression recognition experiments are carried out on facial images without occlusion in this part. All the facial images are divided into six classes, each one corresponding to one of the six facial expressions. We randomly select 1/6 of the samples from each class to form training set. The remaindered 5/6 of the samples per each class is used for testing. After classification, a new subset of 1/6 samples for each class is extracted to form the new training set, and the samples forming the training set are incorporated into the current testing set. This procedure is repeated six times and the average classification accuracies are recorded.

Four different dimensionality reduction methods are used for comparison, namely Eigenfaces, Fisherfces, Laplacianfaces [10], and Sparse Non-negative Matrix Factorization (SNMF) [11]. The recognition accuracies obtained by these methods with SRC (sparse representation classifier) are shown in Fig. 2. It can be seen that SNMF with SRC performs better on whole facial expression images while Laplacianfaces with SRC perform better on facial parts images. Then we compare the classification results of SRC with that of NNC and SVM. The maximum recognition accuracies obtained by these three classifiers are shown in Table 1. The best performances of SRC consistently exceed the best performances of NNC and almost as good as SVM. Then we use voting fusion method and fuzzy integral fusion method to fuse the classification results obtained



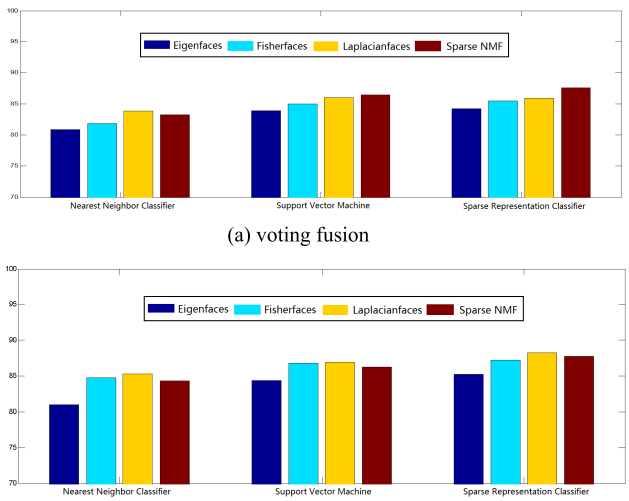
(a) face (b) eye

(c) nose (d) mouth

Fig. 2 Classification results of SRC.

Table 1 Comparison of top recognition rates of NNC, SVM and SRC.

	Face	Eye	Nose	Mouth
Nearest Neighbor Classifier	81.6%	80.4%	79.2%	80.9%
Support Vector Machine	85.6%	82.8%	81.4%	83.0%
Sparse Representation Classifier	86.1%	83.1%	82.5%	84.0%



(a) voting fusion

(b) fuzzy integral fusion

Fig. 3 Comparison of voting fusion and fuzzy integral fusion.

by NNC, SVM, and SRC. The fusing results of the four dimensionality reduction algorithms are illustrated in Fig. 3. SRC with fuzzy integral fusion outperforms other methods. Facial parts based methods take advantages of both holistic representation and local representation, so it obtains better results than using only holistic/local representation.

Fig. 4 Facial expression samples with partial occlusion.

Table 2 Recognition rates for partially occluded images.

Fusion Methods	Eye occlusion	Nose and mouth occlusion
Whole face with NN	79.6%	78.1%
Whole face with SVM	82.1%	81.9%
Whole face with SRC	82.9%	82.3%
Face parts with Voting	84.5%	83.5%
Face parts with Fuzzy Integral	86.6%	85.2%

4.3 Facial Expression Recognition with Partial Occlusion

Some preprocessing is done to get partially occluded facial images. Eye and mouth masks are created to cover the eyes and mouth regions (Some examples are shown in Fig. 4). The maximum recognition accuracies obtained by NNC, SVM, and SRC on the whole face images, and the recognition accuracies obtained using SRC with voting fusion and fuzzy integral fusion are shown in Table 2. We observe that facial expressions are more affected by nose and mouth region than eye region. The classification performance of facial parts based fusing method is better than only using whole facial images for partially occluded facial images. It indicates that our facial parts based sparse representation classifier with fuzzy integral fusing method is efficient for facial expression recognition and it is robust to partial occlusion comparing to other whole face processing methods.

5. Conclusions

Facial parts based sparse representation classification method is proposed for facial expression recognition, and the fusion of multiple classifiers are realized with the aid of fuzzy integral. It has been experimentally demonstrated that the sparse representation classifier got better classification performances than that of NNC and as good as that of SVM. SRC harnesses sparse representation of test sample from each facial expression class, it identifies facial expressions efficiently and it is robust to partial occlusion on facial images.

References

[1] I. Buciu and I. Pitas, Subspace Image Representation for Facial Expression Analysis and Face Recognition and Its Relation to the Human Visual System, Organic Computing, pp.303–320, Springer Verlag, 2008.

[2] M.S. Barlett, G. Littlewort, M. Frank, and C. Lainscsek, "Recognizing facial expression: Machine learning and application to spontaneous behavior," International Conference on Computer Vision and Pattern Recognition (CVPR), pp.568–573, 2005.

[3] R. Duda, P. Hart, and D. Stork, Pattern Classification, 2nd ed. John Wiley & Sons, 2001.

[4] J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman, "Clustering appearances of objects under varying illumination conditions," Proc. IEEE International Conference on Computer Vision and Pattern Recognition, pp.11–18, 2001.

[5] A.Y. Yang, J. Wright, Y. Ma, and S. Sastry, "Feature selection in face recognition: A sparse representation perspective," UC Berkeley Tech Report UCB/EECS-2007-99, 2007.

[6] E. Amaldi and V. Kann, "On the Approximability of minimizing nonzero variables or unsatisfied relations in linear systems," Theoretical Computer Science, vol.209, pp.237–260, 1998.

[7] K. Kwak and W. Pedryca, "Face recognition using fuzzy integral and wavelet decomposition method," IEEE Trans. Syst., Man Cybern. B, Cybern., vol.34, no.4, pp.1666–1675, 2004.

[8] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," IEEE International Conference on Automatic Face and Gesture Recognition, pp.46–53, Grenoble, France, 2000.

[9] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavlets-based facial expression recognition using multi-layer perception," Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp.454–459, 1998.

[10] C. Shan, S. Gong, and P.W. McOwan, "A comprehensive empirical study on linear subspace methods for facial expression analysis," Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), pp.153–158, 2006.

[11] P.O. Hoyer, Non-negative sparse coding. http://www.cs.helsinki.fi/u/phoyer/software.html