ICSP2012 Proceedings

Facial expression recognition based on Gabor wavelets and sparse representation

Shiqing Zhang[1,3] Lemin Li[1] Zhijin Zhao[2]
[1]School of Communication and Information Engineering, University of Electronic Science and Technology of China
Chengdu, China
[2]School of Telecommunication, Hangzhou Dianzi University
Hangzhou, China
[3]School of Physics and Electronic Engineering, Taizhou University
Taizhou, China

Abstract—**The recently-emerged sparse representation in compressive sensing (CS) has gained extensive attention in signal processing and pattern recognition. In this paper, a new method of facial expression recognition based on Gabor wavelets and sparse representation classifier (SRC) is presented. Gabor wavelets representations are firstly extracted to evaluate the performance of the SRC method on facial expression recognition tasks. Three representative classification methods, including artificial neural network (ANN), K-nearest neighbor (KNN), support vector machines (SVM), are used to compare with the SRC method. Experimental results on the popular JAFFE facial expression databases, demonstrate the promising performance of the presented SRC method on facial expression recognition tasks, outperforming the other used methods.**

Keywords- Gabor wavelets; sparse representation; facial expression recognition

I. INTRODUCTION

Facial expressions provide an important behavioral measure for the study of a person's internal emotional states, cognitive processes and social communication. Recently, automatic facial expression recognition has attracted extensive interests in signal processing, computer vision, pattern recognition, and human computer interaction (HCI) research communities. One of the most important applications of facial expression recognition is to make HCI more human-like, more effective, and more efficient [1-3]. Specifically, such computers with the ability of facial expression recognition could detect and track a user's affective states and initiate communications based on this information, rather than simply responding to a user's commands.

Generally, an automatic facial expression recognition system contains two main steps: facial feature extraction, and facial expression classification. The purpose of facial feature extraction is to find the most appropriate representation of the face images for classification. Two well-known approaches for facial feature extraction are geometric feature-based methods and appearance-based methods [4]. The geometric features provide the shape and locations of facial components such as mouth, eyes, brows, noses. The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. The appearance features present the appearance (skin texture) changes of the face, such as wrinkles and furrows. The appearance features can be extracted on either the whole-face or specific regions in a face image. The geometric feature-based methods depend on accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many situations. In contrast, the appearance-based methods, image filters, such as Gabor wavelets, are easily applied to either the whole-face or specific face-regions to extract the appearance changes of a face. Recent reported studies [5, 6] have shown that Gabor wavelet representation is a promising feature extraction technique for facial expression recognition due to its biological relevance and computational properties.

Facial expression classification is to use the extracted facial features to recognize different expressions. Depending on whether the temporal information is considered, facial expression recognition approaches can be categorized as frame-based or sequence-based. The frame-based method does not take the temporal information of input images into account, and use the extracted features from a single image to recognize the expression of that image. In contrast, the sequence-based method attempts to capture the temporal pattern in a sequence to recognize the expression for one or more images. So far, various classifiers, including artificial neural network (ANN) [7], K-nearest neighbor (KNN) [8], support vector machines (SVM) [9], and so on, have been applied for frame-based expression recognition. For sequence-based expression recognition, the widely used techniques are hidden Markov models (HMM) [10], dynamic Bayesian networks [11], and SVM [12].

In recent years, a new theory, Compressive Sensing (CS) [13-15], also referred as Compressed Sensing or Compressive Sampling, has been proposed as a more efficient classification method. The newly-emerged CS theory, originally aiming to address signal sensing and coding problems, claims that a sparse signal can be recovered from a small number of random linear measurements. The CS theory has been used to form a new classification technique called sparse presentation classifier (SRC), showing promising performance on pattern recognition [16-18]. Motivated by very little work done on the

978-1-4673-2197-6/12/$31.00 ©2012 IEEE

application of SRC for facial expression recognition, in this paper we investigate the performance of SRC on facial expression recognition tasks. The representative appearance features, i.e., Gabor wavelets representation, are extracted in our work. And then the state-of-the-art classifiers, including artificial neural network (ANN), K-nearest neighbor (KNN), support vector machines (SVM), are used to compare with SRC. To evaluate the performance of SRC, facial expression recognition experiments are performed on the popular JAFFE database [19].

II. GABOR WAVELETS

The Gabor wavelets representation captures salient visual properties such as spatial localization, orientation selectivity, spatial frequency selectivity, and quadrature phase relationship.

The Gabor wavelets kernel can be defined as

$$\varphi_{\mu,\nu}(z)=\frac{\left\|k_{\mu,\nu}\right\|^2}{\sigma^2}e^{-\frac{\left\|k_{\mu,\nu}\right\|^2\|z\|^2}{2\sigma^2}}[e^{ik_{\mu,\nu}z}-e^{-\frac{\sigma^2}{2}}] \tag{1}$$

where μ and ν denote the orientation and scale of the Gabor kernel, $z=(x,y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector $k_{\mu,\nu}$ is defined as

$$k_{\mu,\nu}=k_\nu e^{i\phi_\mu} \tag{2}$$

where $k_\nu = k_{\max}/f^\nu$ and $\phi_\mu=\pi\mu/8$. $k_{\max}$ is the maximum frequency, and f is the spacing factor between kernels in the frequency domain.

In previous studies [4, 5] one usually use 40 Gabor wavelets at five different scales, $\nu=\{0,1,\cdots,4\}$, and eight orientations, $\mu=\{0,1,\cdots,7\}$, with $\sigma=2\pi, k_{\max}=\pi/2$, and $f=\sqrt{2}$. The Gabor wavelets representation is essentially the concatenated pixels of the 40 modulus-of-convolution images obtained by convolving the input image with these 40 Gabor kernels. In practice, the magnitude of the Gabor wavelets representation is used for facial expression recognition, because they vary slowly with the position while the phases are very sensitive. As suggested in [20], before concatenation each output image is down-sampled by a factor of 16 and normalized to zero mean and unit variance.

III. SPARSE REPRESENTAION

Sparse representation is initially developed as an extension to traditional signal representations such as Fourier and wavelet representations. In recent years, sparse representation has been successfully used to solve many practical problems in signal processing, computer vision, pattern recognition, etc. For instance, in signal and image processing fields, sparse representation is used for signal recovery and acquisition [21], image super-resolution [22], and image sequence denoising [23]. In the emerging field of compressive sensing [13, 15], as a very attractive theory challenging Shannon-Nyquist sampling theorem, sparse representation aims to recover the signal from the compressed measures in a most economical way. Especially, recent researches [16-18] have proved that the classifier based on sparse representation, i.e., SRC, is an exceptionally effective model for classification tasks and yields very promising performance on face recognition.

In SRC, it is assumed that the whole set of training samples form a dictionary, and then the recognition problem is cast as one of discriminatively finding a sparse representation of the test image as a linear combination of training images by solving an optimization problem. Formally, for the training samples of a single class, this assumption can be expressed as

$$\begin{aligned} y_{k,test} &= \alpha_{k,1}y_{k,1}+\alpha_{k,2}y_{k,2}+\cdots+\alpha_{k,n_k}y_{k,n_k}+\varepsilon_k \\ &= \sum_{i=1}^{n_k}\alpha_{k,i}y_{k,i}+\varepsilon_k \end{aligned} \tag{3}$$

where $y_{k,test}$ is the test sample of the k^{th} class, $y_{k,i}$ is the i^{th} training sample of the k^{th} class, $\alpha_{k,i}$ is the weight corresponding weight and ε_k is the approximation error.

For the training samples from all c object classes, the aforementioned Eq.(6) can be expressed as

$$\begin{aligned} y_{k,test} &= \alpha_{1,1}y_{1,1}+\cdots+\alpha_{k,1}y_{k,1}+\cdots \\ &\quad +\alpha_{k,n_k}y_{k,n_k}+\cdots+\alpha_{c,n_c}y_{c,n_c}+\varepsilon \end{aligned} \tag{4}$$

Equivalently,

$$y_{k,test}=\boldsymbol{A}\boldsymbol{\alpha}+\varepsilon \tag{5}$$

where $\begin{cases} A=[y_{1,1}|\cdots|y_{1,n_1}|\cdots|y_{k,1}|\cdots|y_{k,n_k}|\cdots|y_{c,1}|\cdots|y_{c,n_c}] \\ \boldsymbol{\alpha}=[\alpha_{1,1}\cdots\alpha_{1,n_1}\cdots\alpha_{k,1}\cdots\alpha_{k,n_k}\cdots\alpha_{c,1}\cdots\alpha_{c,n_c}]' \end{cases}$.

The linearity assumption in the SRC algorithm coupled with Eq.(5) implies that the weight vector $\boldsymbol{\alpha}$ should be zero except those associated with the correct class of the test sample. To obtain the weight vector $\boldsymbol{\alpha}$, the following l_0 -norm minimization problem should be solved.

$$\min_\alpha\|\boldsymbol{\alpha}\|_0, \quad \text{subject to } \left\|y_{k,test}-\boldsymbol{A\alpha}\right\|_2\le\varepsilon \tag{6}$$

It is known that Eq.(6) is an NP-hard problem. The NP-hard l_0-norm can be replaced by its closest convex surrogate, the l_1-norm. Therefore, the solution of Eq.(6) is equivalent to the following l_1-norm minimization problem.

$$\min_\alpha\|\boldsymbol{\alpha}\|_1, \quad \text{subject to } \left\|y_{k,test}-\boldsymbol{A\alpha}\right\|_2\le\varepsilon \tag{7}$$

This is a convex optimization problem and can be solved by quadratic programming. Once a sparse solution of $\boldsymbol{\alpha}$ is obtained, the classification procedure of SRC is summarized as follows:

Step 1: Solve the l_1-norm minimization problem in Eq.(7).

Step 2: For each class i, compute the residuals between the reconstructed sample $y_{\text{recons}}(i)=\sum_{j=1}^{n_i}\alpha_{i,j}y_{i,j}$ and the given test sample by $r(y_{test},i)=\left\|y_{k,test}-y_{recons}(i)\right\|_2$.

Step 3: The class of the given test sample is determined by identify $(y_{test}) = \arg\min_i r(y_{test}, i)$.

IV. EXPERIMENTS

To illustrate the effectiveness of SRC on facial expression recognition tasks, we firstly extract the Gabor wavelets representations from the original facial images. Subsequently, we conducted facial expression recognition experiments on the popular JAFFE database. The performance of SRC is compared with ANN, KNN and SVM. To reduce the feature length of the Gabor wavelets representations, principal component analysis (PCA) is used for dimensionality reduction. The reduced feature dimension is confined to the range of [0, 100] with an interval of 10.

A 10-fold cross validation scheme is employed in 7-class facial expression recognition experiments, and the average recognition results are reported. In detail, the data sets are split randomly into ten groups of roughly equal numbers of subjects. Nine groups are used as the training data to train a classifier, while the remaining group is used as the testing data. The above process is repeated ten times for each group in turn to be omitted from the training process. Finally, the average recognition results on the testing data are reported.

As a representation ANN, RBFNN is used for its computational simplicity. For the KNN classifier, we set K to be 1 for its satisfying performance. We employ the LIBSVM package, available at http://www.csie.ntu.edu.tw /~cjlin/libsvm, to perform the SVM algorithm with the linear kernel function, one-against-one for multi-class problems. The experiment platform is Intel CPU 2.10 GHz, 1G RAM memory, MATLAB 7.0.1 (R14).

A. *Facial expression database*

The JAFFE database contains 213 images of female facial expressions. Each image has a resolution of 256×256 pixels. The head is almost in frontal pose. The number of images corresponding to each of the seven categories of expressions (anger, joy, sadness, neutral, surprise, disgust and fear) is roughly the same. A few of them are shown in Fig.1.

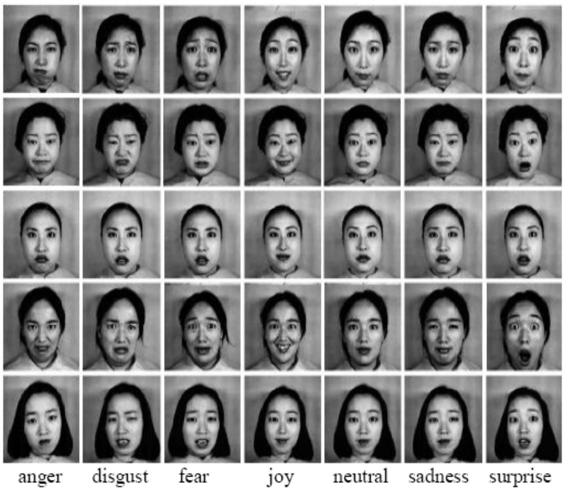

Figure1 Examples of facial expression images from the JAFFE database

B. *Image preprocessing*

Following in [24], we normalized the faces to a fixed distance of 55 pixels between the two eyes. Automatic face registration can be achieved by a robust real-time face detector [25] based on a set of rectangle haar-like features. From the results of automatic face detection, such as face location, face width and face height, two square bounding boxes for left eye and right eye are created respectively. Then, two eyes location can be quickly worked out in terms of the centers of two square bounding boxes for left eye and right eye. Based on the two eyes location, facial images of 110×150 pixels were cropped from original frames. Fig. 2 shows the process of two eyes location and the final cropped image. No further alignment of facial features such as alignment of mouth was performed in our work.

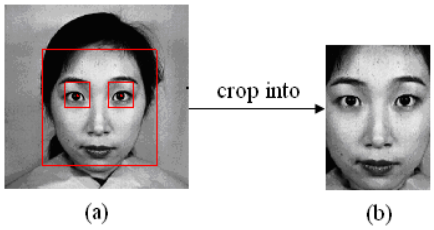

Figure 2. (a) Two eyes location (b) the final cropped image of 110×150 pixels

C. *Experimental results and analysis*

When using the Gabor wavelets representations for facial expression recognition, the recognition results of different classification methods along with reduced dimension of the Gabor wavelets representations are presented in Fig.3. Table I gives the best accuracy of different classification methods with the corresponding reduced dimension of the Gabor wavelets representations. The results in Table I and Fig.3 reveal that SRC achieves an accuracy of 88.57% at best with 60 reduced dimension of the Gabor wavelets representations, outperforming the other used methods. This confirms the validity and high performance of SRC for facial expression recognition.

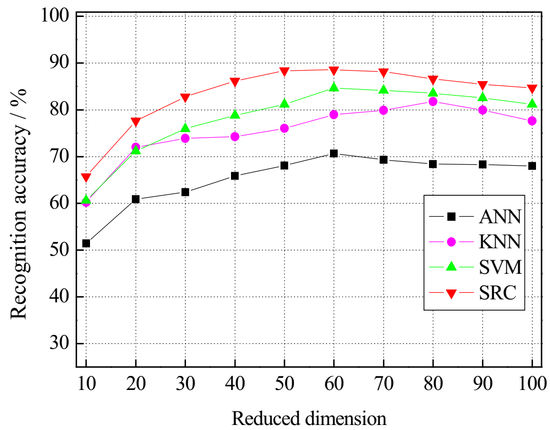

Figure 3 Recognition results on the JAFFE database for different classification methods with the reduced dimension of the Gabor wavelets representations

TABLE I. THE BEST ACCURACY FOR DIFFERENT METHODS

Methods	Dimension	Accuracy (%)
ANN	60	70.64
KNN	80	81.76
SVM	60	84.65
SRC	60	88.57

V. CONCLUSIONS

Automatic facial expression recognition has increasingly attracted attention duo to its important applications to human computer interaction. Designing a good classifier is a crucial step for any successful facial expression recognition system. In this paper, we presented a new method of facial expression recognition via the sparse representation classifier (SRC). The experiment results on the JAFFE database show that the SRC method obtains the promising performance on facial expression recognition tasks due to its good classification property. In our future work, it's an interesting task to employ the SRC technique to develop a real-time facial expression recognition system for natural human-computer interaction.

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, Emotion recognition in human-computer interaction, *IEEE Signal Processing Magazine,* vol. 18, no. 1, pp. 32-80, 2001.

[2] E. Hudlicka, To feel or not to feel: The role of affect in human-computer interaction, *International Journal of Human-Computer Studies,* vol. 59, no. 1-2, pp. 1-32, 2003.

[3] N. Fragopanagos, and J. G. Taylor, Emotion recognition in human-computer interaction, *Neural Networks,* vol. 18, no. 4, pp. 389-405, 2005.

[4] Y. Tian, T. Kanade, and J. F. Cohn, Facial Expression Recognition, *Handbook of face recognition*, pp. 487-519, 2011.

[5] S. Bashyal, and G. Venayagamoorthy, Recognition of facial expressions using Gabor wavelets and learning vector quantization, *Engineering Applications of Artificial Intelligence,* vol. 21, no. 7, pp. 1056-1064, 2008.

[6] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, and H. Lin, Facial expression recognition using radial encoding of local Gabor features and classifier synthesis, *Pattern Recognition,* vol. 45, no. 1, pp. 80-91, 2012.

[7] Y. Tian, T. Kanade, and J. Cohn, Recognizing action units for facial expression analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, no. 2, pp. 97-115, 2001.

[8] N. Sebe, M. S. Lew, Y. Sun, I. Cohen, T. Gevers, and T. S. Huang, Authentic facial expression analysis, *Image and Vision Computing,* vol. 25, no. 12, pp. 1856-1863, 2007.

[9] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, Recognizing facial expression: machine learning and application to spontaneous behavior, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA. pp. 568-573, 2005.

[10] H. Meng, and N. Bianchi-Berthouze, "Naturalistic affective expression classification by a multi-stage approach based on hidden Markov models," *Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science*, pp. 378-387: Springer, 2011.

[11] F. Dornaika, E. Lazkano, and B. Sierra, Improving dynamic facial expression recognition with feature subset selection, *Pattern Recognition Letters,* vol. 32, no. 5, pp. 740-748, 2011.

[12] I. Kotsia, and I. Pitas, Facial expression recognition in image sequences using geometric deformation features and support vector machines, *IEEE Transactions on Image Processing,* vol. 16, no. 1, pp. 172-187, 2007.

[13] D. L. Donoho, Compressed sensing, *IEEE Transactions on Information Theory,* vol. 52, no. 4, pp. 1289-1306, 2006.

[14] R. G. Baraniuk, Compressive sensing [lecture notes], *IEEE Signal Processing Magazine,* vol. 24, no. 4, pp. 118-121, 2007.

[15] E. J. Candes, and M. B. Wakin, An introduction to compressive sampling, *IEEE Signal Processing Magazine,* vol. 25, no. 2, pp. 21-30, 2008.

[16] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, no. 2, pp. 210-227, 2009.

[17] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, Sparse representation for computer vision and pattern recognition, *Proceedings of the IEEE,* vol. 98, no. 6, pp. 1031-1044, 2010.

[18] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, Towards a Practical Face Recognition System: Robust Alignment and Illumination by Sparse Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 99, pp. 1-15, 2011.

[19] M. J. Lyons, J. Budynek, and S. Akamatsu, Automatic classification of single facial images, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, no. 12, pp. 1357-1362, 1999.

[20] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, Classifying facial actions, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, no. 10, pp. 974-989, 1999.

[21] V. Cevher, P. Indyk, L. Carin, and R. Baraniuk, Sparse signal recovery and acquisition with graphical models, *IEEE Signal Processing Magazine,* vol. 27, no. 6, pp. 92-103, 2010.

[22] J. Yang, J. Wright, T. S. Huang, and Y. Ma, Image super-resolution via sparse representation, *IEEE Transactions on Image Processing,* vol. 19, no. 11, pp. 2861-2873, 2010.

[23] M. Protter, and M. Elad, Image sequence denoising via sparse and redundant representations, *IEEE Transactions on Image Processing,* vol. 18, no. 1, pp. 27-35, 2009.

[24] Y. Tian, Evaluation of face resolution for expression analysis, *Proc. of first IEEE Workshop on Face Processing in Video (2004)*, Washington, D.C., USA pp. 82-82, 2004.

[25] P. Viola, and M. Jones, Robust real-time face detection, *International Journal of Computer Vision,* vol. 57, no. 2, pp. 137-154, 2004.