Facial Expression Recognition Using Sparse Representation

SHIQING ZHANG [1], XIAOMING ZHAO [2], BICHENG LEI [1]
[1]School of Physics and Electronic Engineering
Taizhou University
Taizhou 318000
CHINA
tzczsq@163.com, leibicheng@163.com

[2]Department of Computer Science
Taizhou University
Taizhou 318000
CHINA
tzxyzxm@163.com

Abstract: - Facial expression recognition is an interesting and challenging subject in signal processing and artificial intelligence. In this paper, a new method of facial expression recognition based on the sparse representation classifier (SRC) is presented. Two typical appearance facial features, i.e., local binary patterns (LBP) and Gabor wavelets representations are 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 two popular facial expression databases, i.e., the JAFFE database and the Cohn-Kanade database, demonstrate the promising performance of the presented SRC method on facial expression recognition tasks, outperforming the other used methods.
.

Key-Words: - Sparse representation, compressive sensing, facial expression recognition, local binary patterns, Gabor wavelets representations, artificial neural network, K-nearest neighbor, support vector machines

1 Introduction

Emotion conveys the psychological state of a human being, as emotion is expressed by various physiological changes, such as changes in heart beat rate, degree of sweating, blood pressure, and so on. Emotion can be embodied by vocal intonation, facial expression, body gesture and movement. Facial expression is the most natural and efficient means for human beings to communicate their emotions and intentions, as communication is primarily carried out face to face. In recent years, facial expression recognition has attracted increasing attention 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 become 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, a basic automatic facial expression recognition system consists of two steps: facial feature extraction and representation, and facial expression recognition.

Facial feature extraction and representation is to extract facial features to represent the facial changes caused by facial expressions. In facial feature extraction for expression analysis, there are mainly two types of approaches: geometric feature-based methods and appearance-based methods [4]. Geometric features such as the shape and location of the eyes, mouth, brows, nose, etc., are the mostly extracted features. The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. Fiducial facial feature points have been widely adopted as geometric features for facial representation. For example, the geometric positions of 34 fiducial points on a face are usually used to represent facial images [5, 6]. Appearance-based methods work on the facial images directly to represent facial textures such as wrinkles, bulges and furrows, and are thus simple to implement. The representative appearance features contains the raw pixels of facial images, Gabor wavelets representation [7, 8], Eigenfaces [9], and Fisherfaces [10], etc. In recent years, a new face descriptor called local binary patterns (LBP) [11], has been utilized to describe the local appearance of facial expressions with great success on the facial expression recognition tasks [12-14], due to its tolerance against illumination changes and computational simplicity.

Facial expression recognition 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) [15], K-nearest neighbor (KNN) [16], support vector machines (SVM) [17], 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) [18], dynamic Bayesian networks [19], SVM [20].

Among the aforementioned two steps in a basic facial expression recognition system, facial expression recognition is the most critical aspect for any successful facial expression recognition system. The performance of a facial expression recognition system is mainly decided by a classifier. Therefore, designing a good classifier is a crucial step on facial expression recognition tasks.

In recent years, a new theory, Compressive Sensing (CS) [21-23], 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 [24-26]. Motivated by very little work done on the application of SRC for facial expression recognition, in this paper we investigate the performance of SRC on facial expression recognition tasks. Two representative appearance features, i.e., Gabor wavelets representation and local binary patterns (LBP), 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 two popular databases, i.e., the JAFFE database [8] and the Cohn-Kanade database [27].

The remainder of this paper is organized as follows. Compressive sensing (CS) is given in Section 2. Section 3 presents sparse representation classifier (SRC) in detail. In Section 4, two popular facial expression databases are introduced. Facial feature extraction is provided in Section 5. ANN, KNN and SVM is presented in Section 6. Section 7 shows the experiment results and analysis. Finally, the conclusions are given in Section 8.

2 Compressive Sensing

Compressive Sensing (CS) [21-23] is a new sampling strategy that uses a fixed set of linear measurements together with a non-linear recovery process. In order for this scheme to work with a low number of measurements, the CS theory requires the sensed signal to be sparse in a given orthogonal basis and the sensing vectors to be incoherent with this basis.

Mathematically speaking, given a system of under-determined equation:

$$y_{m\times 1} = \mathbf{A}_{m\times n} x_{n\times 1}, \quad m < n \tag{1}$$

It's known that the above Eq.(1) has no unique solution, since the number of variables is larger than the number of equations. In signal processing terms, the length of the signal (n) is larger than the number of samples (m). However, according to the CS theory, if the signal is sparse, it is necessarily unique, and can be reconstructed by practical algorithms.

Suppose that the signal is k-sparse if it is a linear combination of only k basis vectors. That is, there are only k non-zero values in x, and the remainder are all zeroes. In this case, it is possible to find the solution to Eq.(1) by a brute force enumeration of all the possible k-sparse vectors of length n .

Mathematically speaking, this problem can be expressed as

$$\min\|x\|_0, \text{ subject to } y = \mathbf{A}x \quad (2)$$

where $\| \|_0$ is the l_0-norm and denotes the number of non-zero elements in the vector. Eq.(2) is known to be an NP (non-deterministic polynomial) hard problem, and is thus not a practical solution to Eq.(1). The CS literatures [21-23] indicates that under a certain condition on the projection matrix **A**, i.e., restricted isometry property (RIP) [28], the sparsest solution to Eq.(1) can be obtained by replacing the l_0-norm in Eq.(2) by its closest convex surrogate, the l_1-norm ($\| \|_1$). Therefore, the solution to Eq.(2) is equivalent to the following l_1-norm minimization problem

$$\min\|x\|_1, \text{ subject to } y = \mathbf{A}x \quad (3)$$

where the l_1-norm, $\| \|_1$, denotes the minimization of the sum of absolute values of elements in the vector, and serves as an approximation of the l_0-norm.

In practice, the equality $y = \mathbf{A}x$ is often relaxed to take into account the existence of measurement error in the sensing process due to a small amount of noise. Suppose the measurements are inaccurate and consider the noisy model

$$y = \mathbf{A}x + e \quad (4)$$

where e is a stochastic or deterministic error term. Particularly, if the error term e is assumed to be white noise such that $\|e\|_2 < \varepsilon$, where ε is a small constant. A noise robust version of Eq.(3) is defined as follows

$$\min\|x\|_1, \text{ subject to } \|y - \mathbf{A}x\|_2 < \varepsilon \quad (5)$$

3 Sparse Representation Classifier

Based on the CS theory, a new classification method, i.e., sparse representation classifier (SRC), has been recently proposed [24-26]. In the SRC algorithm, 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 the optimization problem in Eq.(3) or (5). Formally, for the training samples of a single class, this assumption can be expressed as

$$\begin{aligned} y_{k,test} &= \alpha_{k,1}y_{k,1} + \alpha_{k,2}y_{k,2} + \cdots + \alpha_{k,n_k}y_{k,n_k} + \varepsilon_k \\ &= \sum_{i=1}^{n_k}\alpha_{k,i}y_{k,i} + \varepsilon_k \end{aligned} \quad (6)$$

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

$$\begin{aligned} y_{k,test} = {} & \alpha_{1,1}y_{1,1} + \cdots + \alpha_{k,1}y_{k,1} + \cdots \\ & + \alpha_{k,n_k}y_{k,n_k} + \cdots + \alpha_{c,n_c}y_{c,n_c} + \varepsilon \end{aligned} \quad (7)$$

Equivalently,

$$y_{k,test} = \mathbf{A}\boldsymbol{\alpha} + \varepsilon \quad (8)$$

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.(8) 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.

$$\min_{\alpha}\|\boldsymbol{\alpha}\|_0, \quad \text{subject to } \|y_{k,test} - \mathbf{A}\boldsymbol{\alpha}\|_2 \le \varepsilon \quad (9)$$

It is known that Eq.(9) 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.(9) is equivalent to the following l_1-norm minimization problem.

$$\min_{\alpha}\|\boldsymbol{\alpha}\|_1, \quad \text{subject to } \|y_{k,test} - \mathbf{A}\boldsymbol{\alpha}\|_2 \le \varepsilon$$
(10)

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.(10).

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$(y_{test}) = \arg\min_i r(y_{test}, i)$.

4 Facial Expression Database

Two popular databases, i.e., the JAFFE database [8] and the Cohn-Kanade database [27], are used to perform facial expression recognition experiments.

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.

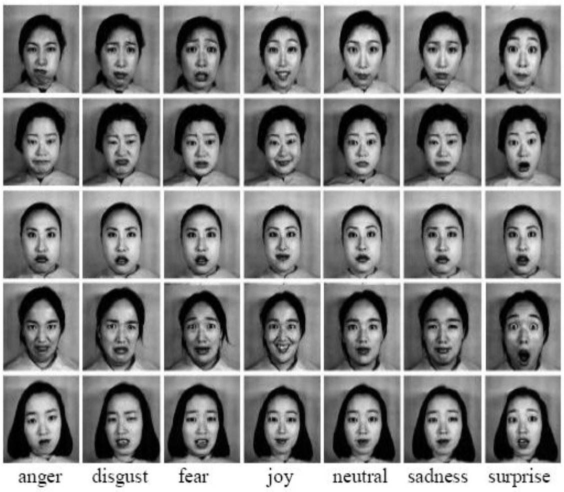

Fig.1 Examples of facial expression images from the JAFFE database

The Cohn-Kanade database consists of 100 university students aged from 18 to 30 years, of which 65% were female, 15% were African-American and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were based on description of prototypic emotions. Image sequences from neutral to target display were digitized into 640×490 pixels with 8-bit precision for grayscale values. Fig.2 shows some sample images from the Cohn-Kanade database. In this work, we selected 320 image sequences from the Cohn-Kanade database. The selected sequences, each of which could be labeled as one of the six basic emotions, come from 96 subjects, with 1 to 6 emotions per subject. For each sequence, the neutral face and one peak frames were used for prototypic expression recognition, resulting in 470 images (32 anger, 100 joy, 55 sadness, 75 surprise, 47 fear, 45 disgust and 116 neutral).

Fig.2 Examples of facial expression images from the Cohn-Kanade database

5 Facial Feature Extraction

Two types of facial features, i.e., Gabor wavelets representation and local binary pattern (LBP), are extracted for facial expression recognition experiments.

5.1 Gabor Wavelets Representation

The Gabor wavelets representation [7, 8] exhibit strong characteristics of spatial locality and orientation selectivity, making them a suitable choice for image feature extraction when one's goal is to derive local and discriminating features for facial expression classification.

The Gabor wavelet kernels can be defined as

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

(11)

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

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \quad (12)$$

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.

As done in [8], we used 40 Gabor wavelet kernels at five 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 Gabor wavelets representation is used for facial expression recognition. As suggested in [29], before concatenation each output image is down-sampled by a factor of 16 and normalized to zero mean and unit variance.

5.2 Local Binary Pattern

The local binary pattern (LBP) operator [11] is a gray-scale invariant texture primitive statistic, which has shown excellent performance in the classification of various kinds of textures. For each pixel in an image, a binary code is produced by thresholding its neighborhood with the value of the center pixel. The LBP code of the center pixel in the neighborhood is obtained by converting the binary code into a decimal one. Based on the LBP operator, each pixel of an image is labeled with an LBP code. The 256-bin histogram of the labels contains the density of each label and can be used as a texture descriptor of the considered region.

The process of LBP features extraction is summarized as follows:

Firstly, a facial image is divided into several non-overlapping blocks. Secondly, LBP histograms are computed for each block. Finally, the block LBP histograms are concatenated into a single vector. As a result, the facial image is represented by the LBP code. Fig.3 presents the process of LBP feature extraction.

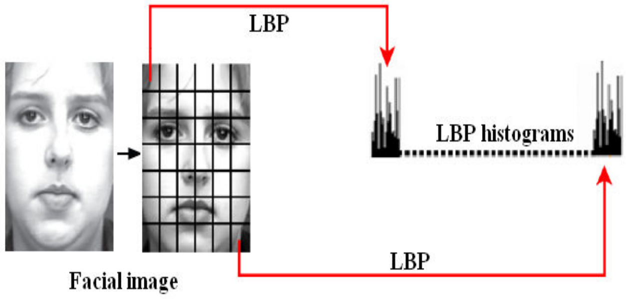

Fig.3 The process of LBP extraction

6 Review of ANN, KNN and SVM

To verify the effectiveness of SRC, three typical classifiers, i.e., artificial neural network (ANN), K-nearest neighbor (KNN), support vector machines (SVM), are employed to compare with SRC. In this section, we separately review ANN, KNN and SVM in brief.

6.1 ANN

Artificial neural network (ANN) is a system derived through models of neuropshychology, and is capable of identifying complex nonlinear relationships between input and output data sets. Generally, ANN can be categorized into three main basic types: multi-layer perceptron (MLP), recurrent neural networks (RNN) and radial basis functions neural networks (RBFNN). In this work, we use RBFNN [30, 31] to perform facial expression recognition, since its main advantages are computational simplicity, supported by well-developed mathematical theory, and robust generalization.

For classification, RBFNN is a three layer feed-forward network that consists of one input layer, one hidden layer and one output layer, as shown in Fig.4. The input layer receives input data. Each input neuron of the input layer corresponds to a component of an input vector $\mathbf{x}$ .The hidden layer is used to cluster the input data and extract features. The hidden layer consists of n neurons and one bias neuron. Each input neuron is fully connected to the hidden layer neurons except the bias one. Each hidden layer neuron computes a kernel function. A Gaussian Radial Basis Function could be a good choice for the hidden layer.

The used Gaussian Radial Basis Function for the hidden layer is defined as

$$y_i = \begin{cases} \exp(-\frac{\|x - c_i\|}{2\sigma_i^2}), & i = 1, 2, \cdots, n \\ 1, & i = 0 \end{cases} \quad (13)$$

where c_i and σ_i represent the center and the width of the neuron, respectively, and the symbol $\| \ \|$ denotes the Euclidean distance. The weight vector between the input layer and the i-th hidden layer neuron corresponds to the center c_i. The closer $\mathbf{x}$ is to c_i, the higher the value the Gaussian function will produce.

The output layer consists of m neurons corresponding to the possible classes of the problem. Each output layer neuron is fully connected to the hidden layer and computes a linear weighted sum of the outputs of the hidden neurons as follows:

$$z_j = \sum_{i=0}^{n} y_i w_{ij}, \quad j = 1, 2, \cdots, m \quad (14)$$

where w_{ij} is the weight between the i-th hidden layer neuron and the j-th output layer neuron.

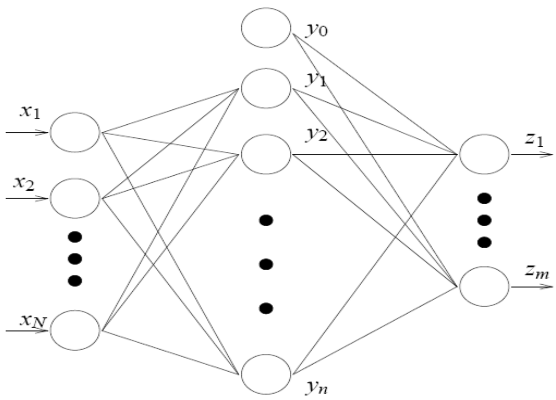

Fig.4 The basic framework of RBFNN

6.2 KNN

K-nearest neighbor (KNN) is an instance-based classification method, firstly introduced by Cover and Hart [32]. KNN has proved popular with facial expression recognition due to its relative simplicity and performance comparable to other methods.

The KNN classification algorithm tries to find the K nearest neighbors of the current sample and uses a majority vote to determine the class label of the current sample. Without prior knowledge, the KNN classifier usually applies the Euclidean distance as the distance metric.

Given two vector $x = (x_1, x_2, \cdots, x_m)$ and $y = (y_1, y_2, \cdots, y_m)$, their Euclidean distance is defined as

$$d(x, y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \quad (15)$$

6.3 SVM

Support vector machines (SVM) is a relatively new machine learning algorithm developed by Vapnik [33]. Based on the statistical learning theory of structural risk management, SVM aims to transform the input vectors to a higher dimensional space by a nonlinear transform, and then an optimal hyperplane which separates the data can be found.

Given the training data set $(x_1, y_1), ..., (x_l, y_l), y_i \in \{-1, 1\}$, to find the optimal hyperplane, a nonlinear transform, $Z = \Phi(x)$, is used to make training data become linearly dividable. A weight w and offset b satisfying the following criteria will be found:

$$\begin{cases} w^T z_i + b \geq 1, & y_i = 1 \\ w^T z_i + b \leq -1, & y_i = -1 \end{cases} \quad (16)$$

The above procedure can be summarized to the following:

$$\min_{w,b} \Phi(w) = \frac{1}{2}(w^T w) \quad (17)$$

subject to $y_i(w^T z_i + b) \geq 1, \quad i = 1, 2, ..., n$

If the sample data is not linearly dividable, the following function should be minimized.

$$\Phi(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \quad (18)$$

whereas ξ can be understood as the error of the classification and C is the penalty parameter for this term.

By using the Lagrange method, the decision function of $w_0 = \sum_{i=1}^{l} \lambda_i y_i z_i$ will be

$$f = \mathrm{sgn}[\sum_{i=0}^{l} \lambda_i y_i (z^T z_i) + b] \quad (19)$$

From the functional theory, a non-negative symmetrical function $K(u,v)$ uniquely defines a Hilbert space H, where K is the rebuild kernel in the space H:

$$K(u,v) = \sum_i \alpha \varphi_i(u) \varphi_i(v) \quad (20)$$

This stands for an internal product of a characteristic space:

$$z_i^T z = \Phi(x_i)^T \Phi(x) = K(x_i, x) \quad (21)$$

Then the decision function can be written as:

$$f = \mathrm{sgn}[\sum_{i=1}^{l} \lambda_i y_i K(x_i, x) + b] \quad (22)$$

The development of a SVM classification model depends on the selection of kernel function. There are four typical kernels that can be used in SVM models. These include linear, polynomial, radial basis function (RBF) and sigmoid function, as described below.

The linear kernel function is defined as

$$K(x_i, x_j) = x_i^T x_j \quad (23)$$

The polynomial kernel function is defined as

$$K(x_i, x_j) = (\gamma x_i^T x_j + coefficient)^{\deg ree} \quad (24)$$

The RBF kernel function is defined as

$$K(x_i, x_j) = \exp(-\gamma \,|\, x_i - x_j \,|^2 \quad (25)$$

The sigmoid kernel function is defined as

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + coefficient) \quad (26)$$

Many real-world data sets involve multi-class problem. Since SVM is inherently binary classifiers, the binary SVM is needed to extend to be multi-class SVM for multi-class problem. Currently, there are two types of approaches for building multi-class SVM. One is the “single machine” approach, which attempts to construct multi-class SVM by solving a single optimization problem. The other is the “divide and conquer” approach, which decomposes the multi-class problem into several binary sub-problems, and builds a standard SVM for each. The most popular decomposing strategy is probably the “one-against-all”. The “one-against-all” approach consists of building one SVM per class and aims to distinguish the samples in a single class from the samples in all remaining classes. Another popular decomposing strategy is the “one-against-one”. The “one-against-one” approach builds one SVM for each pair of classes. When applied to a test point, each classification gives one vote to the winning class and the point is labeled with the class having most votes. In practice, the “one-against-one” approach is more effective than the “one-against-all” approach due to its computation simplicity and comparable performance.

7 Experiments

In this section, we perform facial expression recognition experiments on two popular facial databases, i.e., the JAFFE database and the Cohn-Kanade database, and present experimental results and analysis.

7.1 Experiment setup

For the extraction of the Gabor wavelets representation and LBP, the pre-processing procedure of facial images is given as follows.

Following the setting in [12, 14], we normalized the eye distance of face images to a fixed distance of 55 pixels once the centers of two eyes were located. Generally, it is observed that the width of a face is roughly two times of the distance, and the height is roughly three times. Therefore, based on the normalized value of the eye distance, a resized image of 110 × 150 pixels was cropped from original image. To locate the centers of two eyes, automatic face registration was performed by using a robust real-time face detector based on a set of rectangle Harr-wavelet features [34]. From the results of automatic face detection including face location, face width and face height, two square bounding boxes for left eye and right eye were automatically constructed by using the geometry of a typical up-right face, which has been widely utilized to find a proper spatial arrangement of facial features [35]. Then, the approximate center locations of two eyes can be automatically worked out in terms of the centers of two square bounding boxes for left eye and right eye.

Fig.5 shows the detailed process of two eyes location and the final cropped image from the Cohn-Kanade database. No further alignment of facial features such as alignment of mouth was performed. Additionally, there was no attempt made to remove illumination changes due to LBP's gray-scale invariance.

The cropped facial images of 110 × 150 pixels contain facial main components such as mouth, eyes, brows and noses. The LBP operator is applied to the whole region of the cropped facial images. For better uniform-LBP feature extraction, two parameters, *i.e.*, the LBP operator and the number of regions divided, need to be optimized. Similar to the setting in [12, 14], we selected the 59-bin operator $LBP_{P,R}^{u2}$, and divided the 110 × 150 pixels face images into 18 × 21 pixels regions, giving a good trade-off between recognition performance and feature vector length. Thus face images were divided into 42 (6 × 7) regions, and represented by the LBP histograms with the length of 2478 (59 × 42).

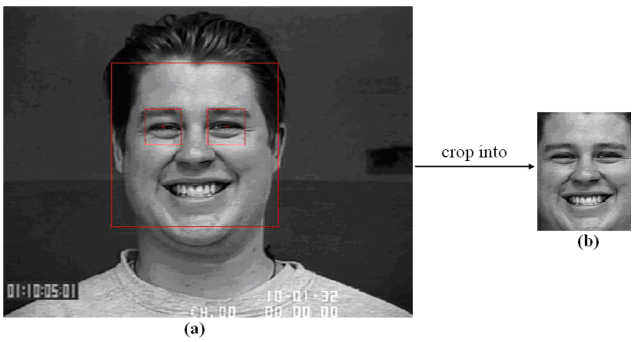

Fig.5 (a) Two eyes location of an original image from the Cohn-Kanade database, (b) The final cropped image of 110 × 150 pixels.

To reduce the feature length of the Gabor wavelets representations, principal component analysis (PCA) [36] 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).

7.2 Experiments on the JAFFE database

When using the LBP features for facial expression recognition, the recognition results of different classification methods on the JAFFE database, including ANN, KNN, SVM and SRC, are given in Table 1. It can be seen from Table 1 that SRC obtains the highest accuracy of 84.76%, outperforming the other used methods. The recognition accuracies using the extracted LBP features for the other used methods, are 68.09% for ANN, 80.95% for KNN and 79.88% for SVM.

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.6. Table 2 gives the best accuracy of different classification methods with the corresponding reduced dimension of the Gabor wavelets representations. The results in Table 2 and Fig.6 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.

Table 1 Comparison of recognition results for different classification methods with the LBP features on the JAFFE database

Methods	Accuracy (%)
ANN	68.09
KNN	80.95
SVM	79.88
SRC	84.76

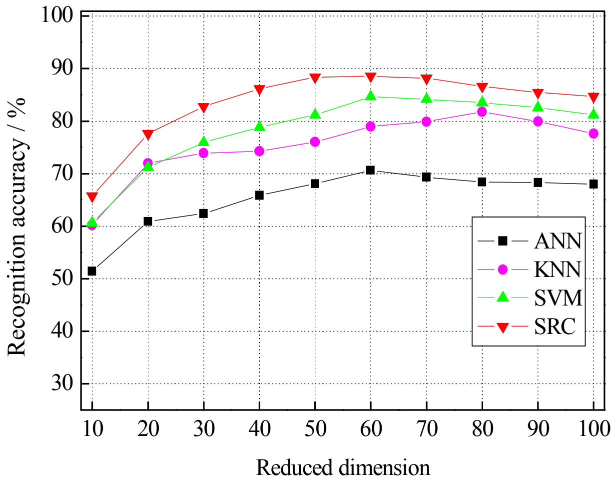

Fig.6 Recognition results on the JAFFE database for different classification methods with the reduced dimension of the Gabor wavelets representations

Table 2 Best results on the JAFFE database for different methods with corresponding reduced dimension of the Gabor wavelets representations

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

To further explore the recognition accuracy per expression, Table 3 and 4 separately present the confusion matrix of 7-class facial expression recognition results with the LBP features and the Gabor wavelets representations. The bold values in Table 3 and 4 represent the recognition accuracy of each expression. From Table 3 and 4, we can observe that three expressions, i.e., anger, joy and neutral, are classified with an accuracy of around 90%, while other four expressions, sad, surprise, disgust and fear, are discriminated with relatively low accuracy (less than 90%).

In our work, the obtained recognition accuracy of SRC (i.e., 84.76% with the LBP features, and 88.57% with the Gabor wavelets representations) for 7-class facial expression recognition on the JAFFE database is highly competitive, compared with previously reported results on the JAFFE database. In [12], similar to our experimental settings, they reported the best accuracy of 81% with SVM and LBP features.

In [37], they extracted the local texture information by applying LBP to facial feature points and obtained an accuracy of 83%.with the nearest neighbour classifier. In [6], by using the Gabor wavelets representations and learning vector quantization (LVQ), they achieved an accuracy of 87.51%. In our previously published work [14], based on LBP and local Fisher discriminant analysis (LFDA), we obtained the best recognition accuracy of 90.7%, outperforming the reported accuracy in this work. Nevertheless, LFDA is used to extract the low-dimensional discriminative embedded data representations from the extracted high-dimensional LBP features with striking performance improvement on facial expression recognition tasks.

7.3 Experiments on the Cohn-Kanade database

Table 5 presents the recognition results of different classification methods with the LBP features on the Cohn-Kanade database. Fig.7 gives the recognition results of different classification methods with the reduced dimension of the Gabor wavelets representations. Table 6 provides the best accuracy for different classification methods with the corresponding reduced dimension.

As shown in Fig.7 and Table 5-6, we can see that SRC still performs best among all used methods for facial expression recognition. In detail, SRC gives an accuracy of 97.14% with the LBP features, and 98.09% with 50 reduced dimension of the Gabor wavelets representations. This indicates the effectiveness of SRC for facial expression recognition, again.

Table 3 Confusion matrix of 7-class facial expression results with the LBP features on the JAFFE database

	Anger (%)	Joy (%)	Sad (%)	Surprise (%)	Disgust (%)	Fear (%)	Neutral (%)
Anger	**93.33**	0	6.67	0	0	0	0

Joy	0	**100**	0	0	0	0	0
Sad	3.22	3.22	**74.19**	3.22	3.22	6.48	6.45
Surprise	0	3.45	3.45	**82.76**	0	10.34	0
Disgust	10.35	0	6.89	0	**82.76**	0	0
Fear	0	0	12.52	3.12	9.37	**71.87**	3.12
Neutral	3.45	0	0	6.89	0	0	**89.66**

Table 4 Confusion matrix of 7-class facial expression results with the Gabor wavelets representations on the JAFFE database

	Anger (%)	Joy (%)	Sad (%)	Surprise (%)	Disgust (%)	Fear (%)	Neutral (%)
Anger	**96.55**	0	3.45	0	0	0	0
Joy	0	**90.32**	9.68	0	0	0	0
Sad	6.45	3.22	**87.10**	0	3.23	0	0
Surprise	0	3.45	0	**86.21**	0	6.89	3.45
Disgust	7.14	0	0	0	**82.14**	10.72	0
Fear	0	0	9.37	6.25	6.25	**78.13**	0
Neutral	0	0	0	0	0	0	**100**

Table 5 Comparison of recognition results for different classification methods with the LBP features on the Cohn-Kanade database

Methods	Accuracy (%)
ANN	93.45
KNN	96.22
SVM	95.24
SRC	97.14

Table 6 Best results on the Cohn-Kanade database for different methods with corresponding reduced dimension of the Gabor wavelets representations

Methods	Dimension	Accuracy (%)
ANN	70	94.86
KNN	60	97.14
SVM	80	96.16
SRC	50	98.09

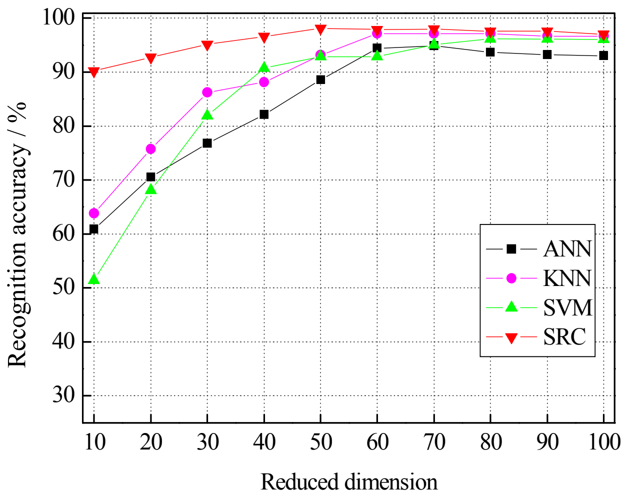

Fig.7 Recognition results on the Cohn-Kanade database for different classification methods with the reduced dimension of the Gabor wavelets representations

Table 7 and 8 presents the confusion matrix of 7-class expression recognition results with the LBP features and the Gabor wavelets representations, respectively. As shown in Table 7-8, it can be seen that most of seven expressions are identified very well with an accuracy of 100%.

Now we compare our reported results (i.e., 97.14% with the LBP features, and 98.09% with the Gabor wavelets representations) with previously reported results on the Cohn-Kanade database. In [12], they obtained a 7-class recognition accuracy of 91.4% at best with LBP features and SVM. In [38], they obtained the highest accuracy of 93.4% with SVM on 7-class tasks, but they used an improved LBP features called local directional pattern (LDP).

Table 7 Confusion matrix of 7-class facial expression results with the LBP features on the Cohn-Kanade database

	Anger (%)	Joy (%)	Sad (%)	Surprise (%)	Disgust (%)	Fear (%)	Neutral (%)
Anger	**90**	0	0	0	0	0	10
Joy	0	**100**	0	0	0	0	0
Sad	3.33	0	**90**	0	0	0	6.67
Surprise	0	0	0	**100**	0	0	0
Disgust	0	0	0	0	**100**	0	0
Fear	0	0	0	0	0	**100**	0
Neutral	0	0	0	0	0	0	**100**

Table 8 Confusion matrix of 7-class facial expression results with the Gabor wavelets representations on the Cohn-Kanade database

	Anger (%)	Joy (%)	Sad (%)	Surprise (%)	Disgust (%)	Fear (%)	Neutral (%)
Anger	**100**	0	0	0	0	0	0
Joy	0	**100**	0	0	0	0	0
Sad	0	0	**100**	0	0	0	0
Surprise	0	0	0	**96.67**	0	0	3.33

Disgust	10	0	0	0	**90**	0	0
Fear	0	0	0	0	0	**100**	0
Neutral	0	0	0	0	0	0	**100**

8 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 and the Cohn-Kanade 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. In addition, it's also interesting to investigate the performance of the SRC technique to predict the behavior of financial time series [39].

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant No.61203257 and No.61272261, and Zhejiang Provincial Natural Science Foundation of China under Grant No.Z1101048 and No.Y1111058.

References:

[1] R Cowie, E Douglas-Cowie, N Tsapatsoulis *et al.*, Emotion recognition in human-computer interaction, *IEEE Signal Processing Magazine,* Vol. 18, No. 1, 2001, pp. 32-80.

[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, 2003, pp. 1-32.

[3] M Kurematsu, J Hakura, and H Fujita, A Framework of a Speech Communication System with Emotion Processing, *WSEAS Transactions on Systems and Control*, Vol. 3, No. 6, 2007, pp. 283-290.

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

[5] W Zheng, X Zhou, C Zou *et al.*, Facial expression recognition using kernel canonical correlation analysis (KCCA), *IEEE Transactions on Neural Networks,* Vol. 17, No. 1, 2006, pp. 233-238.

[6] 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, 2008, pp. 1056-1064.

[7] B Kara, N Watsuji, Using Wavelets for Texture Classification, *WSEAS Transactions on Computers*, 2003, pp. 920-924.

[8] 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, 1999, pp. 1357-1362.

[9] M A Turk, and A P Pentland, Face recognition using eigenfaces, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1991, pp. 586-591.

[10] P N Belhumeur, J P Hespanha, and D J Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 19, No. 7, 1997, pp. 711-720.

[11] T Ojala, M Pietik inen, and T M enp, Multiresolution gray scale and rotation invariant texture analysis with local binary patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 24, No. 7, 2002, pp. 971-987.

[12] C Shan, S Gong, and P McOwan, Facial expression recognition based on Local Binary Patterns: A comprehensive study, *Image and Vision Computing,* Vol. 27, No. 6, 2009, pp. 803-816.

[13] X Zhao, S Zhang, Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding, *EURASIP Journal on Advances in Signal Processing*, 2012, 20.

[14] S Zhang, X Zhao, and B Lei, Facial Expression Recognition Based on Local Binary Patterns and Local Fisher Discriminant Analysis,

WSEAS TRANSACTIONS on Signal Processing, Vol. 8, No. 1, 2012, pp. 21-31.

[15] 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, 2001, pp. 97-115.

[16] N Sebe, M S Lew, Y Sun *et al.*, Authentic facial expression analysis, *Image and Vision Computing,* Vol. 25, No. 12, 2007, pp. 1856-1863.

[17] M Bartlett, G Littlewort, M Frank *et al.*, Recognizing facial expression: machine learning and application to spontaneous behavior, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 568-573.

[18] 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.

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

[20] 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, 2007, pp. 172-187.

[21] D L Donoho, Compressed sensing, *IEEE Transactions on Information Theory,* Vol. 52, No. 4, 2006, pp. 1289-1306.

[22] R G Baraniuk, Compressive sensing [lecture notes], *IEEE Signal Processing Magazine,* Vol. 24, No. 4, 2007, pp. 118-121.

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

[24] J Wright, A Y Yang, A Ganesh *et al.*, Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 31, No. 2, 2009, pp. 210-227.

[25] J Wright, Y Ma, J Mairal *et al.*, Sparse representation for computer vision and pattern recognition, *Proceedings of the IEEE,* Vol. 98, No. 6, 2010, pp. 1031-1044.

[26] A Wagner, J Wright, A Ganesh *et al.*, Towards a Practical Face Recognition System: Robust Alignment and Illumination by Sparse Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, No. 99, 2011, pp. 1-15.

[27] T Kanade, Y Tian, and J Cohn, Comprehensive database for facial expression analysis, Proc. International Conference on Face and Gesture Recognition, 2000, pp. 46-53.

[28] E J Candes, and T Tao, Decoding by linear programming, *IEEE Transactions on Information Theory,* Vol. 51, No. 12, 2005, pp. 4203-4215.

[29] G Donato, M Bartlett, J Hager *et al.*, Classifying facial actions, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 21, No. 10, 1999, pp. 974-989.

[30] J Park, and I W Sandberg, Universal approximation using radial-basis-function networks, *Neural computation,* Vol. 3, No. 2, 1991, pp. 246-257.

[31] S Zhang, X Zhao, B Lei, Spoken emotion recognition using radial basis function neural network, CSEE2011-Part 1, *Communications in Computer and Information Science (CCIS)*, Springer, 214, 2011, pp. 437-442.

[32] T Cover, and P Hart, Nearest neighbor pattern classification, *IEEE Transactions on Information Theory,* Vol. 13, No. 1, 1967, pp. 21-27.

[33] V Vapnik, *The nature of statistical learning theory*: Springer-Verlag, New.York, 2000.

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

[35] P Campadelli, R Lanzarotti, G Lipori *et al.*, Face and facial feature localization, Proc. International Conference on Image Analysis and Processing, 2005, pp. 1002-1009.

[36] I T Jolliffe, *Principal component analysis*, Second edition ed., New York: Springer, 1986.

[37] X Feng, B Lv, Z Li *et al.*, A Novel Feature Extraction Method for Facial Expression Recognition, Proc. Joint Conference on Information Sciences, 2006, pp.

[38] T Jabid, M H Kabir, and O Chae, Robust Facial Expression Recognition Based on Local Directional Pattern, *ETRI journal,* Vol. 32, No. 5, 2010, pp. 784-794.

[39] F Neri, Agent Based Modeling Under Partial and Full Knowledge Learning Settings to Simulate Financial Markets", *AI Communications*, 25 (4), IOS Press, 2012, pp. 295-305.