Neural Comput & Applic
DOI 10.1007/s00521-014-1569-1

A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering

Ayşegül Uçar · Yakup Demir · Cüneyt Güzeliş

Received: 12 September 2013 / Accepted: 11 March 2014

Abstract In this paper, a novel algorithm is proposed for facial expression recognition by integrating curvelet transform and online sequential extreme learning machine (OSELM) with radial basis function (RBF) hidden node having optimal network architecture. In the proposed algorithm, the curvelet transform is firstly applied to each region of the face image divided into local regions instead of whole face image to reduce the curvelet coefficients too huge to classify. Feature set is then generated by calculating the entropy, the standard deviation and the mean of curvelet coefficients of each region. Finally, spherical clustering (SC) method is employed to the feature set to automatically determine the optimal hidden node number and RBF hidden node parameters of OSELM by aim of increasing classification accuracy and reducing the required time to select the hidden node number. So, the learning machine is called as OSELM-SC. It is constructed two groups of experiments: The aim of the first one is to evaluate the classification performance of OSELM-SC on the benchmark datasets, i.e., image segment, satellite image and DNA. The second one is to test the performance of the proposed facial expression recognition algorithm on the Japanese Female Facial Expression database and the Cohn-Kanade database. The obtained experimental results are compared against the state-of-the-art methods. The results demonstrate that the proposed algorithm can produce effective facial expression features and exhibit good recognition accuracy and robustness.

Keywords Online sequential extreme learning machine · Local curvelet transform · Spherical clustering · Facial expression recognition

1 Introduction

During the last few decades, research community has witnessed considerable interest in facial expression recognition due to its imperative applications in different areas such as human–computer interaction, human–robot interaction, forensic, medical treatment, virtual reality and data-driven animation [1–4]. According to the Facial Action Coding System (FACS) in [5], the facial expressions basically consist of six class, i.e., happiness, sadness, fear, anger, disgust and surprise. On the other hand, the humans have special facial expressions relating to only themselves. Each of the facial expression groups can be represented by all kinds of facial expressions. Since the face recognition under facial expression variation and the facial expression recognition from the images including different facial expressions are still challenging problems and are being done researches for solution [6, 7].

Extracting the representative facial features from the face images is the best important step at facial expression recognition. In the literature, the facial expression recognition algorithms were constructed based on feature-based methods extracting the shape and location information of

A. Uçar (✉)
Mechatronics Engineering Department, Firat University, 23119 Elazig, Turkey
e-mail: agulucar@firat.edu.tr

Y. Demir
Electrical-Electronics Engineering Department, Firat University, 23119 Elazig, Turkey
e-mail: ydemir@firat.edu.tr

C. Güzeliş
Electrical-Electronics Engineering Department, Izmir University of Economics, 35330 Balcova, Izmir, Turkey
e-mail: cuneyt.guzelis@ieu.edu.tr

Published online: 25 March 2014

significant regions of the face geometry [8, 9] or appearance-based methods extracting facial features by a set of filters [10–13]. In this paper, the curvelet transform is used for extracting the image features. Curvelet transform is an efficient feature extraction tool achieving localization in both time domain and frequency domain [14]. A lot of curves and lines of a face can easily be captured better than wavelet transform concerning the point singularities in a face [14–17]. However, it is generally used together with the dimension reduction techniques since the curvelet transform extracts very large features similar to the other appearance-based methods.

In this paper, it is applied the curvelet transform to local subregions of image to get rid of above disadvantage. Some statistical features relating to these local regions are calculated. In the literature, there are available some local feature extraction methods such as Local Binary Patterns (LBP) [18] and Scale Invariant Feature Transformation (SIFT) [19] operating on certain small blocks of whole face image to improve the recognition performance. In the proposed facial expression recognition algorithm, the distinctive feature sets generated by local curvelet transform are used to train and test a classifier. The training process and storing space of a classifier increase in a case of the large training data. This is not appropriate for online applications. In order to take action on the difficult more effectively, the different online learning algorithms such as resource allocation network (RAN) [20], minimal RAN (MRAN) [21], growing and pruning radial basis function (GAP-RBF) [22] and generalized adaptive resonance theory (GART) [23] were proposed. However, their accustomed insufficiencies are that too many parameters need to be determined and the training time is too long. In this paper, it is introduced a new version of online sequential extreme learning machine (OSELM) proposed by Huang et al. for classifying discriminative facial features by reducing these limitations [24, 25]. OSELM on the basis of extreme learning machines (ELM) provides one-by-one and chunk-by-chunk updating and computationally efficiency [26]. ELM is generalization of single-hidden layer feed-forward networks (SLFNs) [26, 27]. Differences between ELM and SLFN are that the hidden layer parameters of ELMs do not need to iteratively tune, the weights of hidden nodes and biases are randomly chosen, output weights are analytically determined [26, 28, 29] and ELM can be trained in much faster speed and produces better generalization performance [30]. However, OSELM needs to determine the hidden node number similar to ELM. In [31], it was introduced structure-adjustable algorithm to the training stage of OSELM. This algorithm called as SAO-ELM was resulted in large hidden node number although it dispensed the parameter selection time. In this paper, the hidden node number and the hidden layer parameters of OSELM with RBF hidden nodes are automatically determined by spherical clustering (SC) method in [32]. The obtained centers and radius of the spheres are transformed to the parameters of RBF hidden nodes. It is referred as OSELM-SC. Thus, the generalization performance of OSELM is improved thanks to the optimal hidden node parameters. Moreover, it is used less hidden node than OSELM and does not spare extra time and struggle to determine their number.

The proposed novel facial expression recognition algorithm consists of the following steps: (i) the face in image is detected and cropped, (ii) the cropped image is preprocessed in order to provide further illumination invariant, (iii) the curvelet transform is applied to the local regions of the face image, (iv) the statistical features relating to the regions are extracted and (v) OSELM-SC classifier is applied to the obtained features.

This paper is organized as follows. Section 2 describes the learning algorithms of OSELM and OSELM-SC. Section 3 presents a short review of curvelet transform. Section 4 introduces the proposed algorithm for facial expression recognition. Section 5 gives comparatively experimental results for both OSELM-SC classifiers and the proposed facial expression recognition algorithm. Section 6 provides conclusions.

2 Online sequential extreme learning machine and its version based on spherical clustering

This section reviews the OSELM and OSELM-SC classifiers.

2.1 Online sequential extreme learning machine with radial basis function

Given a training set consisting of N samples, $(x_j, t_j) \in \mathcal{R}^n x \mathcal{R}^m$, where x_j is an input vector and t_j is the associated target vector, an OSELM with M hidden nodes is formulated as

$$s(x_j) = \sum_{m=1}^{M} \beta_m G(c_m, b_m, x_j) = t_j \quad j = 1, \ldots, N, \tag{1}$$

where $G(c_m, b_m, x_j)$ shows the output of the m-th hidden node with radial basis activation function (RBF) for the input x_j described by

$$G(c_m, b_m, x_j) = \exp\left\{-b_m \|x_j - c_m\|^2\right\}, \quad b_m \in \mathcal{R}^+ \tag{2}$$

(c_m, b_m) refers to the learning parameters of hidden nodes and β_m is the weight vector connecting the m-th hidden node to the output nodes.

OSELM is trained at two stages: The initialization stage and sequential learning stage [24]. In the initialization stage, only N_0 training samples are used to train the OSELM with the condition $N_0 \geq M$. The following steps are applied:

1. All input weights c_m and b_m are given randomly.
2. The initial hidden layer output matrix, H_0, and the initial output weights, β^0, are calculated, respectively:

$$H_0 = \begin{bmatrix} G(c_1, b_1, x_1) & \ldots & G(c_M, b_M, x_1) \\ \vdots & \ldots & \vdots \\ G(c_1, b_1, x_{N_0}) & \ldots & G(c_M, b_M, x_{N_0}) \end{bmatrix}_{N_0 x M}, \tag{3}$$

$$\mathbf{P}_0 = \left(\mathbf{H}_0^T \mathbf{H}_0\right)^{-1}. \tag{4}$$

$$\beta^0 = \mathbf{P}_0 \mathbf{H}_0^T \mathbf{T}_0. \tag{5}$$

where $\mathbf{T}_0 = [\mathbf{t}_1, \ldots, \mathbf{t}_{N_0}]$.

3. k is set to 0 and then the sequential learning stage is run.

The sequential learning stage is completed at three steps:

1. The $(k+1)$-th chunk of data is taken.
2. The new partial hidden layer output matrix H_{k+1} and the output weights β^{k+1} are calculated by

$$H_{k+1} = [G(c_1, b_1, x_1) \ldots G(c_M, b_M, x_{k+1})] \tag{6}$$

$$P_{k+1} = P_k - P_k H_{k+1}^T \left(I + H_{k+1} P_k H_{k+1}^T\right)^{-1} H_{k+1} P_k \tag{7}$$

$$\beta^{k+1} = \beta^k + P_{k+1} H_{k+1}^T \left(T_{k+1} - H_{k+1} \beta^k\right) \tag{8}$$

where $\mathbf{T}_{k+1} = [\mathbf{t}_1\, \mathbf{t}_{k+1}]$.

3. k is set to $k+1$ and it is run the first step of sequential learning stage. After $k = N$, the training of OSELM is stopped and the OSELM output $s(\mathbf{z})$ assigned to an unknown input vector $\mathbf{z}$ is calculated by c_m, b_m, and the final value of β_m.

$$s(\mathbf{z}) = \sum_{m=1}^{M} \beta_m G(\mathbf{c}_m, \mathbf{b}_m, \mathbf{z}). \tag{9}$$

2.2 Online sequential extreme learning machine with radial basis function initialized by spherical clustering

Given $\left(x_j, t_m^j\right)$ with $x_j \in \mathcal{R}^n$, $t_m^j \in \{-1, 1\}$ for $j = 1, \ldots, N$ and $m = 1, \ldots, M$ defines a set of N samples each of which is associated with one of the M classes $(M \geq 1)$ such that any label $t_m^j = 1$ for a unique m and -1 for the others [32]. M-class spherical classes are defined by

$$S_m = \{x \in \mathcal{R}^n \left\| (x_j - c_m) \right\|^2 - R_m^2 = 0, \quad j = 1, \ldots, N, \quad m = 1, \ldots, M\} \tag{10}$$

where $R_m \in \Re$ and $c_m \in R^n$ are the radius and center of the surface S_m, respectively. S_m sphere includes the associated samples, and the misclassification error is zero if it is satisfied the following quadratic inequalities,

$$\left\| (x_j - c_m) \right\|^2 - R_m^2 > 0, \quad t_m^j = -1 \tag{11}$$

$$\left\| (x_j - c_m) \right\|^2 - R_m^2 \leq 0, \quad t_m^j = 1 \tag{12}$$

To get the smallest volume robust spheres including the associated class samples, the optimization problem is constructed as

$$\min_{c_m, R_m} \quad g(R_m) = \sum_{m=1}^{M} R_m^2 \tag{13}$$

$$\text{subject to} \quad \left\| (x_j - c_m) \right\|^2 - R_m^2 > 0, \quad t_m^j = -1 \tag{14}$$

$$\left\| (x_j - c_m) \right\|^2 - R_m^2 \leq 0, \quad t_m^j = 1 \tag{15}$$

To obtain a solution in a case of that the samples can not be spherically separated, the constrained optimization problem in Eq. (13) can be transformed using the penalty function method into an unconstrained one:

$$\min_{c_m, R_m} \quad L = \sum_{m=1}^{M} \sum_{j=1}^{N} f\left(t_m^j \left(\left\| (x_j - c_m) \right\|^2 - R_m^2 \right)\right) + A \sum_{m=1}^{M} R_m^2 \tag{16}$$

where $A > 0$ is the regularization parameter that should be chosen sufficiently small to include almost all class samples by the spheres but sufficiently large for preventing over-fitting and for excluding noise and outliers, and the penalty function $f(\cdot)$ is chosen as the ramp function:

$$f(\xi) = \xi \quad \xi > 0 \quad f'(\xi) = 1 \quad \xi > 0$$
$$f(\xi) = 0 \quad \xi \leq 0 \quad f'(\xi) = 0 \quad \xi < 0 \tag{17}$$

The optimization problem in Eq. (16) is solved by the gradient-descent method with adaptive learning rate. The negative gradients of L cost function with respect to c_m centers and R_m radii are calculated:

$$-\frac{\partial L}{\partial c_m} = 2 \sum_{j=1}^{N} t_m^j (x_j - c_m) f'\left(t_m^j \left(\left\| (x_j - c_m) \right\|^2 - R_m^2 \right)\right) \tag{18}$$

$$-\frac{\partial L}{\partial R_m} = 2 \sum_{j=1}^{N} t_m^j f'\left(t_m^j \left(\left\| (x_j - c_m) \right\|^2 - R_m^2 \right)\right) R_m - 2AR_m \tag{19}$$

c_m centers and R_m radii are updated as follows.

$$\Delta c_m(\upsilon+1) = \alpha \Delta c_m(\upsilon) - (1-\alpha)\eta(\upsilon)\frac{\partial L}{\partial c_m} \tag{20}$$

$$\Delta R_m(\upsilon+1) = \alpha \Delta R_m(\upsilon) - (1-\alpha)\eta(\upsilon)\frac{\partial L}{\partial R_m} \tag{21}$$

where α is the momentum constant and $\eta(\upsilon)$ is the adaptive learning rate for υ-th epoch. The parameters are conventionally determined as in [33, 34]. On the other hand, the regularization parameter A is determined adaptively as follows.

if $A < N$,

if $L(\upsilon+1) < L(\upsilon)$, $A^{\upsilon+1} = 0.01 + A^{\upsilon}$

elseif $L(\upsilon+1) > L(\upsilon)$, $A^{\upsilon+1} = A^{\upsilon} - 0.01A^{\upsilon}$ end

end

In the beginning, each class is separately clustered by k—means. The cluster number is automatically adjusted taking into consideration the quantization error ratio. All radius and center parameters are then updated by the proposed unconstrained optimization problem for only 30 epochs.

In order to determine the RBF hidden node parameters of OSELM-SC, the obtained final center and radius values are used. The center values of OSELM-SC are set to c_m, and b_m values are calculated by

$$b_m = 1/\left(2 \times R_m^2\right). \tag{22}$$

The output of OSELM-SC is expressed as

$$s(x_j) = \sum_{m=1}^{M} \beta_m \exp\left\{-1/\left(2 \times R_m^2\right)\left\|x_j - c_m\right\|^2\right\} = t_j \quad j = 1, \ldots, N. \tag{23}$$

3 Curvelet transform

Fourier transform is far from being universally effective in image processing. Many researchers have sought alternatives to Fourier transform and consequently introduced new tools such as Gabor transform, wavelet transform and curvelet transform. Curvelet transform is a multi-scale and multi-directional tool. It was first proposed based on an anisotropic geometric wavelet transform, called as ridgelet transform in [35]. It has high directional sensitivity and anisotropy. The curvelet transform allows almost optimal sparse representation of curve singularities because of its parabolic scaling properties [35]. In the transform, the image is firstly smooth partitioned to analyze local line or curve singularities. The curve singularities in each smooth partitioned block are then approximated to a line singularity. The ridgelet transform is finally applied to the obtained blocks, subimages. This block ridgelet-based transform was referred as curvelet transform. Because of more time-consuming and unclear of the geometry of ridgelet geometry, second generation curvelet transform was proposed based on the Fourier transform in order to handle image boundaries in [36]. The curvelet transform is faster, less complex, and less redundant as compare to the other transforms [36].

Curvelet transform is defined by the inner product

$$c(j,l,k) := \left\langle f, \varphi_{j,l,k} \right\rangle, \tag{24}$$

where $\varphi_{j,l,k}$ is the curvelet basis functions, f is signal and j, l, and k are scale and orientation, and position, respectively.

The continuous curvelet transform is represented by r and θ polar coordinates in the frequency domain. Figure 1 shows how the image is decomposed by the radial divisions into multiple scales and the angular divisions into different positions or orientation. For each $j \geq j_0$, the gray window in Fig. 1 is defined as

$$U_j(r,\theta) = 2^{-\frac{3j}{4}} W(2^{-j}r) V\left(\frac{2^{\left[\frac{j}{2}\right]}\theta}{2\pi}\right), \tag{25}$$

where $[j/2]$ is the integer part of $j/2$, and $W(r)$ and $V(t)$ are the windows used for radial and angular partition, respectively. The windows obey to the basic conditions

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1 \quad r \in \left(\frac{3}{4}, \frac{3}{2}\right), \tag{26}$$

$$\sum_{l=-\infty}^{\infty} V^2(t-l) = 1 \quad t \in \left(-\frac{1}{2}, \frac{1}{2}\right). \tag{27}$$

They are smooth nonnegative and real valued on $t \in [-11]$ and $r \in [1/2, 1]$.

Curvelet are defined as function of $x = (x_1, x_2)$ by rotating and shifting φ_j at the scale 2^{-j} by

$$\varphi_{j,l,k}(x) = \varphi_j\left(R_{\theta_l}\left(x - x_k^{(j,l)}\right)\right), \tag{28}$$

where orientation $\theta_l = 2\pi 2^{\left[-\frac{j}{2}\right]} l$, $l = 0, 1, \ldots,$ with $0 \leq \theta_l \leq 2\pi$, position $x_k^{(j,l)} = x R_{\theta_l}^{-1}(2^{-j}k_1, 2^{-j/2}k_2)$, shift

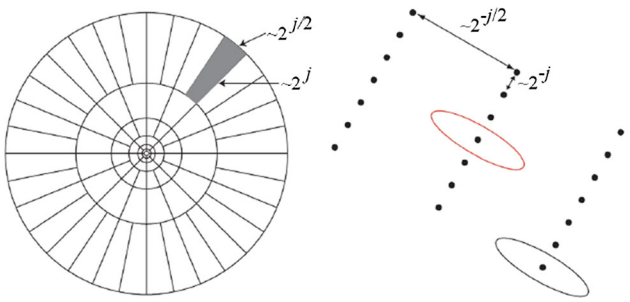

Fig. 1 Space–frequency tiling in curvelet domain [36]

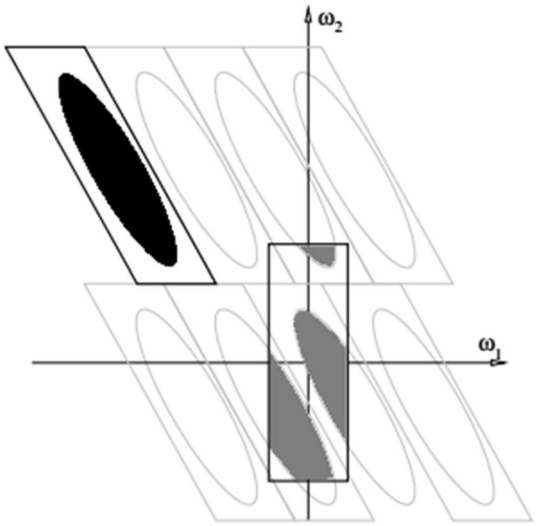

Fig. 2 Wrapping a segment around the origin [36]

parameter $k = (k_1, k_2) \in \mathbb{Z}^2$, and R_{θ_l} is an orthogonal rotation in radians.

The curvelet coefficients are expressed as an integral over the entire frequency plane by using Plancherel's theorem.

$$c(j,l,k) := \frac{1}{2\pi^2}\int \hat{f}(\omega)\overline{\varphi_{j,l,k}(x)}\mathrm{d}\omega = \frac{1}{2\pi^2}\int \hat{f}(\omega)U_j(R_{\theta_l}\omega)\mathrm{e}^{\mathrm{j}x_k^{(i,j)},\omega}\mathrm{d}\omega. \quad (29)$$

Let the input $f[t_1,\ t_2]$ $(0 \cdot t_1,\ t_2 < n)$ in the spatial Cartesian, then the discrete curvelet transform can be defined by

$$c^D(j,l,k) := \sum_{0 \le t_1,t_2 < n} f[t_1,t_2]. \quad (30)$$

The curvelet transform based on wrapping technique is applied at four steps:

1. Compute 2D Fast Fourier Transform (FFT) coefficients,
2. Generate a product of U_j for each scale and angle,
3. Index the data by wrap this product around the origin as shown in Fig. 2,
4. Apply a 2D inverse FFT and collect the discrete curvelet coefficients in spatial domain.

4 Proposed facial expression recognition algorithm

In this section, it is introduced the proposed facial expression recognition algorithm. The proposed algorithm relies on image decomposition by curvelet transform for facial expression recognition and the performance of OSELM-SC classifiers. A block schematic diagram of the proposed facial expression recognition algorithm is shown in Fig. 3. The algorithm is applied at sixfolds:

1. Detect the face regions both in the training and testing images by Viola-Jones algorithm [37],
2. Apply histogram equalization to all images in order to obtain further robustness against illumination [38],
3. Divide the whole face images into local regions and apply curvelet transform to the regions,
4. Extract some statistical features, i.e., standard deviation, mean and entropy,
5. Spherically separate a part of training data and determine the hidden node number and the parameters of the hidden node,
6. Apply OSELM-SC classifiers to the qualified coefficients.

Figure 4 shows the some curvelet coefficients at the first three scales relating to the curvelet transform with four scales and eight different angles applied to an image from the JApanese Female Facial Expression database (JAFFE) [39]. As seen from Fig. 4, the approximation accuracy of face image reduces as the scale number increases. In this paper, the curvelet coefficients of only first scale were used.

5 Experimental results

In this section, two different experiments are constructed to evaluate the performance of the proposed algorithm. The first experiment is built to show the performance of OSELM-SC classifiers on benchmark classification problems and to compare them with the other online learning algorithms in the literature, such as OSELM, GART, MRAN and GAP-RBF. On the other hand, the second experiment is arranged to test the proposed new face expression recognition algorithm on the JAFFE database [39] and the Cohn-Kanade database [40]. The comparative experiments with the facial expression algorithms in the literature are carried out. In addition, the advantages of OSELM-SC are presented on a facial expression recognition problem. All the experiments are conducted in MATLAB 7 environment on a computer with 3.4-GHz core-i7-CPU and Windows 7 operation system.

5.1 Classification by benchmark datasets

In the first experiment, the performance of OSELM-SC is evaluated on three benchmark classification problems, i.e., image segment, satellite image and DNA from the University of California at irvine repository of the machine learning database [41]. The properties of the problem set are described in Table 1. In the image segmentation

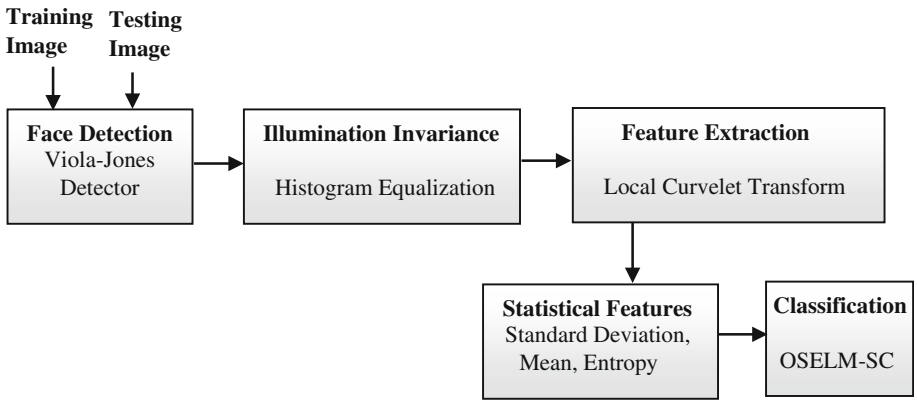

Fig. 3 The block schema of the proposed facial expression recognition algorithm

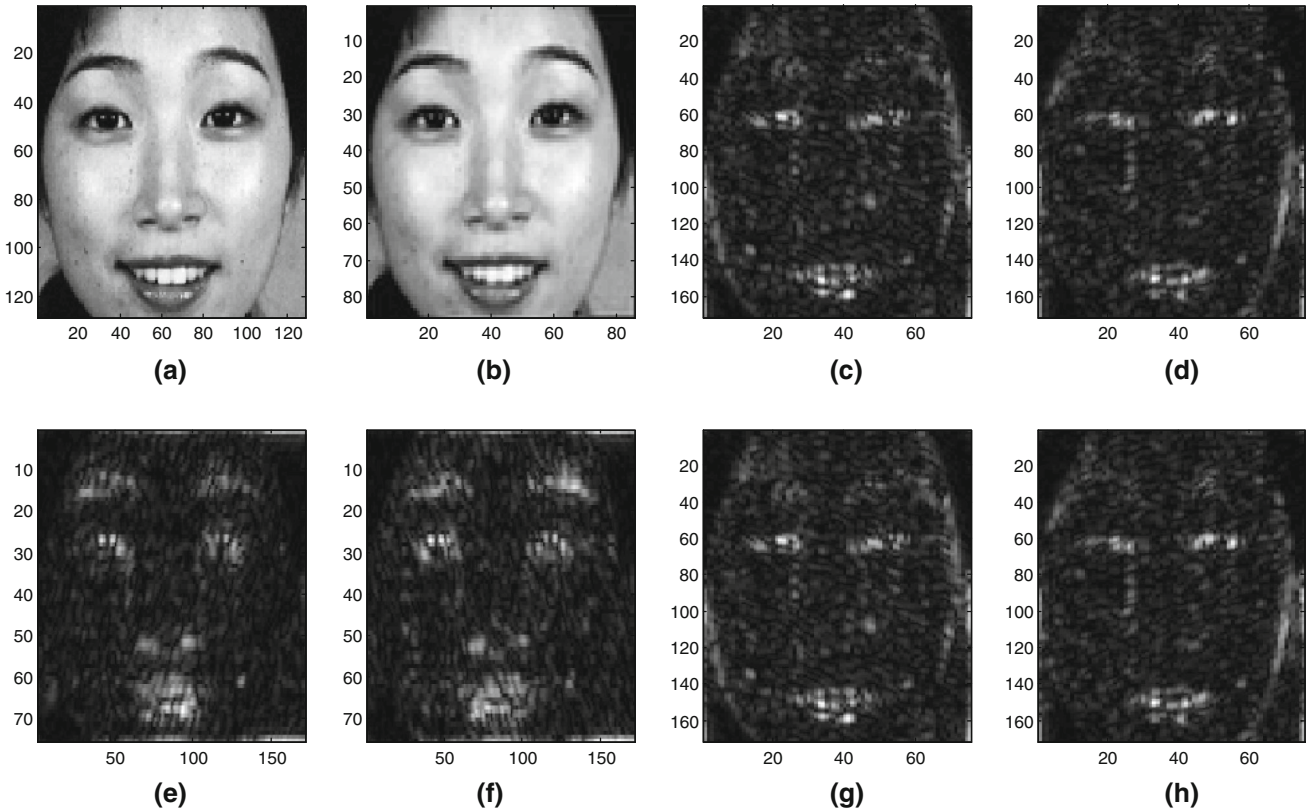

Fig. 4 **a** Cropped original JAFFE image, **b** curvelet subband at the first scale, **c**, **d** curvelet subbands at the second scale and the first two angular orientations and **e**–**h** curvelet subbands at the third scale and the first four angular orientations

Table 1 Characteristics of benchmark classification problems

Problem	# Input space	# Class	# Training data	# Testing data
Image segment	19	7	1,500	810
Satellite image	36	6	4,435	2,000
DNA	180	3	2,000	1,186

problem, each image was separated into the regions with 3×3 pixels. The total number of regions in database is 2,310. In this classification problem, each region is assigned to any one of seven image parts, i.e., brick facing, sky, foliage, cement, window, path and grass. Nineteen Different features relating to each region are used as inputs. The training and testing data sets are randomly selected from the database. In the satellite image problem, one frame of landsat multispectral scanner imagery is composed of four digital images of the same scene in four different spectral bands. Each scene image having 82×100 pixels was divided into the regions with 3×3 pixels. In this classification problem, the central pixel in each region is assigned to the six categories, i.e., red soil,

cotton crop, gray soil, damp gray soil, soil with vegetation stubble and very damp gray soil. Thirty-six spectral features extracted of each region are used as inputs. The DNA (primate splice-junction gene sequences) problem consists of recognizing the boundaries between exons (the regions of the DNA sequence retained after splicing) and introns (the regions of the DNA sequence that are spliced out) in a DNA sequence. The boundaries are classified into three groups as: exon/intron (EI sites), intron/exon (IE sites) and neither (sites). In the problem, a DNA sequence includes 60 elements called as "nucleotides" or "base-pairs." In the problem, the symbolic variables in the DNA sequence are shown by three binary indicator variables; 180 binary features are employed as inputs. The training and test sets of DNA and satellite image problems were previously fixed in [41], but the order of training set was randomly mixed up for each trial as in [24] in this paper.

The experimental results are also compared with several related methods, namely GART, MRAN, GAP-RBF and SAO-ELM. For the simulations of OSELM, it should previously be determined only the optimal number of hidden nodes. In this paper, the hidden node number of OSELM was determined by using twofold cross-validation method as in [24]. The training set was divided into two nonoverlapped sets as the training and validation ones. The number of hidden nodes is gradually increased to $0.1 \times N$ with an interval of five. The optimal hidden node number is selected one exhibiting the averaged smallest validation error over ten trials. The number of initial training samples (N_0) is set to be equal to $N_0 = M + 100$ as stated in [24]. In Table 2, it is comparatively summarized the hidden node number, the parameter selection time, the average training and testing times, and the average training and testing classification accuracy. The parameter selection means the duration passing through ten trials. OSELM-SC is run ten times in order to compare with OSELM in the same circumstance in that parameter selection time although the proposed OSELM-SC algorithm is not more sensitive to the initial parameters thanks to SC algorithm, and just one trial is absolutely sufficient for SC. On the other hand, the results relating to classification accuracy were averaged over fifty trials for OSELM similar to [24] and over ten trials for OSELM-SC. The results relating to other networks were cited directly from the literature [21–24, 31]. In addition, the training, testing and parameter selection times of GART, MRAN and GAP-RBF are not given in Table 2 due to the use of the computers with different properties in [21–23, 31]. However, it can be controlled from [24] and [31] that they have seven more times training times at least over OSELM [24]. The training time of SAO-ELM is less in a small rate than that of OSELM in [31].

As seen from Table 2, OSELM-SC has better generalization performance than OSELM and exhibits also good performance with respect to the results directly taken from the literature. In addition, it can be deduced that OSELM-SC is better than OSELM in all aspects including the parameter selection time, training time, testing time, training accuracy and testing accuracy. All of these advantages come with small hidden node number and

Table 2 Recognition rates of the proposed algorithm on benchmark classification problems

Problem	Algorithms	# Hidden node number	Training accuracy	Testing accuracy	Parameter selection time (s)	Training time (s)	Testing time (s)
Image segment	OSELM-SC	133	97.00	94.57	48.48	0.1248	0.0156
	OSELM [24]	180	96.18	94.10	49.06	0.1947	0.0172
	MRAN [21]	53.1	–	93.30	–	[a]	–
	GAP-RBF [22]	44.2	–	89.93	–	[a]	–
	GART [23]	–	98.04	90.53	–	–	–
	SAO-ELM [31]	191	97.25	95.16	–	–	–
Satellite image	OSELM-SC	233	94.88	90.55	50.33	0.5928	0.1248
	OSELM [24]	400	93.83	89.25	200.51	1.8861	0.2455
	MRAN [21]	20.4	–	86.36	–	[a]	–
	GART [23]	–	98.04	90.53	–	–	–
	SAO-ELM [31]	413	94.80	91.39	–	–	–
DNA	OSELM-SC	3	95.05	93.34	14.83	0.0780	0.000
	OSELM [24]	200	95.32	93.26	770.98	2.0567	0.6019
	MRAN [21]	5	–	86.85	–	[a]	–
	GART [23]	–	1.00	88.31	–	–	–
	SAO-ELM [31]	214	96.40	94.54	–	–	–

[a] The values are at least seven more times than those of OSELM in [24, 31]

Fig. 5 Facial expression samples of the JAFFE database

optimal parameters of RBF hidden node; especially the results relating to DNA problem shows drastically the performance of OSELM-SC with three hidden nodes.

5.2 Facial expression recognition experiment

In the second experiment, the proposed facial expression recognition algorithm is evaluated on the JAFFE database [39] and the Cohn-Kanade database [40]. The JAFFE database includes 213 grayscale facial expression images relating to 10 Japanese females [39]. There are seven different facial expressions consisting of happy, sadness, fear, surprise, angry, disgust and neutral. Each subject has two to four different images with the resolution of 256 × 256 pixels for each expression. Figure 5 shows some images comprising seven basic facial expressions from the JAFFE database. In this paper, the images of JAFFE are posed as seven-class classification problem.

Cohn-Kanade database consists of 2,105 images of 182 subjects in the age range of 18–30 years [40]. Images have the resolution of 640 × 480 or 640 × 490 pixels. About 65 % of subjects are female, 15 % is African-American and 3 % is Asian or Latino. All of the image sequences start with a neutral expression and finish at the expression apex. The discriminative images of an expression are located in the last frames of the sequences. In this paper, the images of Cohn-Kanade database are posed as six-class classification problem. Figure 6 shows some samples of six basic facial expressions from the Cohn-Kanade database. Cohn-Kanade database was commonly used in the literature [40]. However, it was used different image sequences in there. It is difficult to find same sequences and same labeling in there. In this paper, it was chosen 1,184 images including one of the six expressions, four images for each expression of 92 subjects are taken in accordance with [42] in order to make a fair comparison. The images are selected from the last frame of each sequence.

In the facial expression algorithm, after being detected the face locations in all images by Viola-Jones algorithm, they were cropped and scaled to 128 × 128 pixels resolution. To get the illumination invariance, the histogram equalization method was applied to all cropped images. The images were divided into the regions of 16 × 16 pixels. The curvelet transform was applied to the images at four scales and eight different angles to the local regions. Statistical features of relating to only the first scale were calculated since the first scale provided an efficient correctness [17, 34]. The Euclidean norms of all approximate

Fig. 6 Facial expression samples of the Cohn-Kanade database

coefficients were used due to their complex value [34]. The proposed OSELM-SC classifiers were used for classifying the statistical features.

In order to evaluate the performance of the proposed facial expression algorithm in this paper, leave-one-subject-out cross-validation method was used [12, 18, 42–47]. In the leave-one-subject-out evaluation method, the test set was composed of images of one subject, out of the 10 total and the remaining images were used to generate the training set. The training set does not contain any image of the test subject at each trial. Since the test image is not included in training set, this method is more appropriate and consistent than leave-one-out evaluation method to show the generalization ability of algorithm. In the JAFEE database, the images of 10 subjects in the database are separated into 10 sets, each of which includes images of one subject. All images in the Cohn-Kanade database were separated into 10 sets, and all images of one subject were included in the same set.

The performance of OSELM-SC in the proposed algorithm was compared to k-NN with 1-Nearest Neighbor, ELM and OSELM. Tables 3 and 4 show the performances of the proposed algorithm on the JAFEE and Cohn-Kanade databases. From the tables, it can be observed that OSELM-SC has the highest testing accuracy, the smallest hidden node number and the lowest parameter learning time. Furthermore, it can be inferred that the proposed algorithm extracts the efficient features for recognizing facial expressions.

The confusion matrix analyzes the confusion between the expressions. Tables 5 and 6 show the averaged confusion matrix obtained by using OSELM-SC on the JAFFE and the Cohn-Kanade databases, respectively. The confusion matrix explains that all of the expressions are recognized with high accuracy and that the biggest confusion takes place on the fear, disgust, and angry expressions. On JAFFE database, the fear expression is confused with the disgust 6.10 %, the surprise 1.70 %, the sadness 1.64 % and the angry 0.46, whereas the angry expression is confused with the disgust 1.60 %, the fear 3.00 % and the sadness 2.65 %, and the disgust expression is confused with the fear 5.79 % and the happy 0.91 %. Once again, on the Cohn-Kanade database, it is observed that the largest difficulties occur on the classification of the disgust and anger expressions. Explicitly, the disgust expression is misclassified as the fear 5.06 % and the anger 2.14 %, whereas the angry expression is misclassified as the disgust 4.55 % and the fear 1.90 %. This situation explains that each person shows the disgust, angry and fear expressions in a lot different ways from the others. The happiness, surprise and sadness expressions may be expressed in a similar way.

Tables 7 and 8 demonstrate the comparison with the reported results of the algorithms in [12, 42–47] for the JAFFE database and in [18, 42] for the Cohn-Kanade database. These algorithms were selected because of available the state-of-the-art performance using the same evaluation method and the same databases. However, the recognition results in [44] were produced by taking away two JAFFE images called as KR.SR3.79 and NA.SU1.79. As observed in Tables 7 and 8, the proposed approach is better than all seven algorithms in [12, 42–47] for the JAFFE database and three algorithms in [18, 42] for the Cohn-Kanade database in view of average recognition rate. The recognition rate of the proposed algorithm by using the OSELM-SC was calculated to be 94.65 and 95.17 % by leave-one-subject-out evaluation method on the JAFFE and

Table 3 Recognition rates for different classifiers of the proposed algorithm on the JAFFE database

Classifier	Training accuracy	Testing accuracy	Parameter selection time (s)	Hidden node number	Training time (s)	Testing time (s)
OSELM-SC	94.45	94.65	20.53	7	0.015	0.0
ELM	95.45	94.44	74.13	128	0.046	0.0
OSELM	94.45	94.56	64.53	128	0.109	0.0
k-NN	–	86.66	–	–	–	0.031

Table 4 Recognition rates for different classifiers of the proposed algorithm on the Cohn-Kanade database

Classifier	Training accuracy	Testing accuracy	Parameter selection time (s)	Hidden node number	Training time (s)	Testing time (s)
OSELM-SC	95.65	95.15	22.49	6	0.015	0.0
ELM	94.45	94.64	85.18	148	0.046	0.031
OSELM	96.26	94.86	79.10	148	0.124	0.031
k-NN	–	85.33	–	–	–	0.026

Table 5 Recognition rates of the proposed algorithm for the JAFFE database

	Happy	Sadness	Fear	Surprise	Angry	Disgust	Neutral
Happy	**98.55**	0.0	0.0	0.0	0.0	1.20	0.25
Sadness	0.0	**94.25**	0.0	0.61	3.13	0.0	2.01
Fear	0.0	1.64	**90.10**	1.70	0.46	6.10	0.0
Surprise	0.0	0.0	3.00	**96.20**	0.0	0.80	0.0
Angry	0.0	2.65	3.00	0.0	**92.75**	1.60	0.0
Disgust	0.91	0.0	5.79	0.0	0.0	**93.30**	0.0
Neutral	0.0	1.32	3.63	0.0	1.35	0.0	**93.70**

The bold values indicate the best recognition rate

Table 6 Recognition rates of the proposed algorithm for the Cohn-Kanade database

	Happy	Sadness	Fear	Surprise	Angry	Disgust
Happy	**98.40**	0.0	0.0	0.0	0.0	1.60
Sadness	0.0	**97.00**	0.0	1.20	1.80	0.0
Fear	0.0	1.64	**94.20**	2.70	1.46	0.0
Surprise	0.0	0.0	3.40	**96.00**	0.0	0.60
Angry	0.0	0.0	1.90	0.0	**93.55**	4.55
Disgust	2.14	0.0	5.06	0.0	0.0	**92.80**

The bold values indicate the best recognition rate

Table 7 Comparison with state-of-the-art performance of the proposed algorithm on the JAFFE database

Study	Method	Recognition rate (%)
The proposed	Local curvelet transform	94.65
[12], 2012	Radial encoded Gabor jets	89.67
[42], 2011	Patch-based-Gabor	91.00
[43], 2010	Salient feature vectors	85.92
[44], 2008	WMMC	65.77
[45], 2007	KCCA	67.00
[46], 2008	DCT	79.30
[47], 2010	FEETS + PRNN	83.84

Table 8 Comparison with state-of-the-art performance of the proposed algorithm on the Cohn-Kanade database

Study	Method	Recognition rate (%)
The proposed	Local curvelet transform	95.17
[18], 2009	Boosted LBP	95.15
[18], 2009	LBP	92.60
[42], 2011	Patch-based-Gabor	94.48

Cohn-Kanade databases, respectively. The recognition rate is 4.98 and 3.65 percent higher than those in [12, 42], respectively, on JAFFE database, as well as 2.57 and 0.69 percent higher than those in [18, 42], respectively, on Cohn-Kanade database.

6 Conclusions

In this paper, an efficient and robust algorithm for recognizing facial expressions was developed by using OSELM-SC and curvelet transform. The proposed algorithm has the following four useful features: (1) a high generalization performance and a fast training speed of OSELM, (2) the capability of extracting the effective facial expression features including many curves and lines of curvelet transform, (3) the property of reducing the large curvelet coefficients calculating by reduced statistical features of local regions and (4) the efficiency of automatic and fast determination of the optimal hidden layer parameters and optimal hidden node number of OSELM by SC method.

In the proposed algorithm, Viola-Jones algorithm is firstly used to detect the face regions in images, and then, the histogram equalization is applied to the cropped face images to achieve more illumination invariance. The curvelet transform is then applied to each local region of images, and the statistical features such as standard deviation, means and entropy are extracted. Finally, SC method is applied to the statistical features. The obtained centers and radius values are used to calculate the hidden layer parameters of OSELM. OSELM-SC classifier with optimal parameters is employed to recognize the facial expressions.

A comparative study was conducted on a set of online learning methods such as MRAN, GAP-RBF, OSELM and SAO-ELM on benchmark classification problems. The experimental results showed that OSELM-SC classifier significantly surpassed the others and OSELM in terms of the smallest hidden node number, the shortest parameter selection time, highest recognition performance and small testing times.

Meanwhile, the obtained features by proposed facial expression recognition algorithm were classified by using OSELM-SC, OSELM, ELM and k-NN. Experimental results demonstrated that it was extracted efficient features by the proposed algorithm, and OSELM-SC performed significantly better in recognition of facial expressions in terms of recognition performance and testing speed. It was

observed that the algorithm achieved the testing accuracy (94.65 %) for JAFFE and the testing accuracy (95.17 %) for Cohn-Kanade. Furthermore, in the comparisons with the state-of-the-art methods in facial expression recognition proved that the proposed algorithm was superior to other popular algorithms for facial expression recognition at testing accuracy.

In addition, the proposed approach can also be applied to many applications, i.e., patient state detection, driver fatigue monitoring and intelligent tutoring system. In a possible future work, the proposed algorithm may be extended to obtain more robustness on significant face regions including eyes, eyebrows and lips.

References

1. Anderson K, McOwan PW (2006) A real-time automated system for recognition of human facial expressions. IEEE Trans Syst Man Cybern B Cybern 36:96–105
2. Cohn I, Sebe N, Chen L, Garg A, Huang TS (2003) Facial expression recognition from video sequences: temporal and static modeling. Comput Vis Image Underst 91(1–2):160–187
3. Zhang Y, Ji Q (2003) Facial expression understanding in image sequences using dynamic and active visual information fusion. IEEE International Conference on Computer Vision, vol 2. Nice, France, pp 113–118
4. Tian YL, Kanade T, Cohn JF (2005) In: Li SZ, Jain AK (eds) Handbook of face recognition. Springer, Heidelberg, pp 247–276
5. Ekman P, Friesen WV (1971) Constant across cultures in the face and emotion. J Pers Soc Psychol 17(2):124–129
6. Lajevardi SM, Hussain ZM (2010) Higher order orthogonal moments for invariant facial expression recognition. Digit Signal Process 20(6):1771–1779
7. Lu X, Wang Y, Jain AK (2003) Combining classifiers for face recognition. Int Conf Multimed Expo 3:13–16
8. Donato G, Bartlett M, Hager J, Ekman P, Sejnowski T (1999) Classifying facial actions. IEEE Trans Pattern Anal Mach Intell 21(10):974–989
9. Tian YL, Kanade T, Cohn J (2001) Recognizing action units for facial expression analysis. IEEE Trans Pattern Anal Mach Intell 23(2):97–115
10. Zhen W, Huang TS (2003) Capturing subtle facial motions in 3d face tracking. Ninth IEEE Int Conf Comput Vis, vol 2. Nice, France, pp 1343–1350
11. Lyons M, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with Gabor wavelets. In: IEEE 3rd international conference on automatic face and gesture recognition, pp 200–205
12. Gu W, Xiang C, Venkatesh YV, Huang D, Lin H (2012) Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. Pattern Recognit 45(1):80–91
13. Uçar A (2013) Facial expression recognition based on significant face components using steerable pyramid transform. In: International conference on image processing, computer vision and pattern recognition, vol 2. Las Vegas, USA, pp 687–692
14. Candes EJ, Donoho DL (2000) Curvelets—a surprisingly effective nonadaptive representation for objects with edges. Vanderbilt University Press, Nashville, TN
15. Mandal T, Wu JQM, Yuan Y (2009) Curvelet based face recognition via dimensional reduction. Signal Process 89(12):2345–2353
16. Mohammed AA, Minhas R, Wu JQM, Sid-Ahmed MA (2011) Human face recognition based on multidimensional pca and extreme learning machine. Pattern Recognit 44(10–11):2588–2597
17. Uçar A (2012) Color face recognition based on curvelet transform. In: International conference on image processing, computer vision and pattern recognition, vol 2. Las Vegas, USA, pp 561–566
18. Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis Comput 27(6):803–816
19. Li ZS, Imai J, Kaneko M (2010) Facial expression recognition using facial–component–based bag of words and PHOG descriptors. J Inst Image Inform Telev En 64(2):230–236
20. Platt JA (1991) Resource-allocating network for function interpolation. Neural Comput 3(2):213–225
21. Lu Y, Sundararajan N, Saratchandran PA (1997) Sequential learning scheme for function approximation using minimal radial basis function (RBF) neural networks. Neural Comput 9(2):461–478
22. Huang GB, Saratchandran P, Sundararajan N (2004) An efficient sequential learning algorithm for growing and pruning RBF (GAP-RBF) networks. IEEE Trans Syst Man Cybern 34(6):2284–2292
23. Yap KS, Lim CP, Abidin IZ (2008) A hybrid ART-GRNN online learning neural network with a ε-insensitive loss function. IEEE Trans Neural Netw 19(9):1641–1646
24. Liang NY, Huang GB, Saratchandran P, Sundararajan N (2006) A fast and accurate online sequential learning algorithm for feedforward network. IEEE Trans Neural Netw 17(6):1411–1423
25. Lan Y, Soh YC, Huang GB (2009) Ensemble of online sequential extreme learning machine. Neurocomputing 72(13–15):3391–3395
26. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
27. Demir Y, Uçar A (2003) Modelling and simulation with neural and fuzzy-neural networks of switched circuits. COMPEL: Int J Comput Math Electr Electro Eng 22(2):253–272
28. Choi K, Toh KA, Byun H (2012) Incremental face recognition for large-scale social network services. Pattern Recognit 45(8):2868–2883
29. Uçar A (2014) Color face recognition based on steerable pyramid transform and extreme learning machines. Sci World J 2014:1–45. Article Id: 628494
30. Uçar A, Yavşan E (2014) Behavior learning of memristor—based chaotic circuit by extreme learning machines. Turk J Elec Eng Comp Sci. doi:10.3906/elk-1304-248
31. Li G, Liu M, Dong M (2010) A new online learning algorithm for structure-adjustable extreme learning machine. Comput Math Appl 60(3):377–389
32. Uçar A, Demir Y, Güzeliş C (2014) A penalty function method for designing efficient robust classifiers with input–space optimal separating surfaces. Turk J Elec Eng Comp Sci. doi:10.3906/elk-1301-190
33. Haykin S (2008) Neural networks and learning machines, 3rd edn. Prentice Hall, New Jersey, USA
34. Uçar A, Demir Y, Güzeliş C (2006) A new formulation for classification by ellipsoids. In: Savacı FA (ed) TAINN 2005. LNAI, vol 3949. Springer, Heidelberg, pp 100–106
35. Do MN, Vetterli M (2003) The finite ridgelet transform for image representation. IEEE Trans Image Process 12(1):16–28
36. Candes EJ, Demanet L, Donoho DL, Ying L (2006) Fast Discrete curvelet transforms. Multiscale Model Simul 5(3):861–899
37. Viola P, Jones M (2004) Robust real–time face detection. Int J Comput Vision 57(2):137–154

38. Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, Romeny BH, Zimmerman JB, Zuiderveld K (1987) Adaptive histogram equalization and its variations. Comp Vis Graph 39(3):355–368
39. Lyons M, Budynek J, Akamatsu S (1999) Automatic classification of single facial images. IEEE Trans Pattern Anal Mach Intell 21(12):1357–1362
40. Kanade T, Cohn JF, Yingli T (2000) Comprehensive database for facial expression analysis. In: IEEE 4th international conference on automatic face and gesture recognition. Pittsburgh, PA, USA, pp 46–53
41. Blake CL, Merz CJ (1998) UCI Repository of machine learning databases. http://archive.ics.uci.edu/ml/datasets.html, Department of Information and Computer Science, University of California, Irvine
42. Zhang L, Tjondronegoro D (2011) Facial expression recognition using facial movement features. IEEE Trans Affect Compt 2(4):219–229
43. Kyperountas M, Tefas A, Pitas I (2010) Salient feature and reliable classifier selection for facial expression classification. Pattern Recognit 43(3):972–986
44. Zhengdong C, Bin S, Xiang F, Yu-Jin Z (2008) Automatic coefficient selection in weighted maximum margin criterion. In: 19th International conference on pattern recognition. Tampa, FL, pp 1–4
45. Horikawa Y (2007) Facial expression recognition using KCCA with combining correlation kernels and Kansei information. In: International conference on computational science and its applications. Kuala Lampur, Malaysian, pp 489–498
46. Bin J, Guo-Sheng Y, Huan-Long Z (2008) Comparative study of dimension reduction and recognition algorithms of DCT and 2DPCA. In: International conference on machine learning and cybernetics. Kunming, China, pp 407–410
47. Wong J, Cho SA (2010) Face emotion tree structure representation with probabilistic recursive neural network modeling. Neural Comput Appl 19(1):33–54