

RESEARCH

Open Access

Gauss–Laguerre wavelet textural feature fusion with geometrical information for facial expression identification

Ahmad Poursaberi^{1*}, Hossein Ahmadi Noubari², Marina Gavrilova¹ and Svetlana N Yanushkevich¹

Abstract

Facial expressions are a valuable source of information that accompanies facial biometrics. Early detection of physiological and psycho-emotional data from facial expressions is linked to the situational awareness module of any advanced biometric system for personal state re/identification. In this article, a new method that utilizes both texture and geometric information of facial fiducial points is presented. We investigate Gauss–Laguerre wavelets, which have rich frequency extraction capabilities, to extract texture information of various facial expressions. Rotation invariance and the multiscale approach of these wavelets make the feature extraction robust. Moreover, geometric positions of fiducial points provide valuable information for upper/lower face action units. The combination of these two types of features is used for facial expression classification. The performance of this system has been validated on three public databases: the JAFFE, the Cohn-Kanade, and the MMI image.

Keywords: Facial expression, Gauss–Laguerre wavelet, Feature fusion, Texture analysis

Introduction

Automatic facial expression recognition (AFER) is of interest to researchers because of its importance for facial biometric-based intelligent support systems. It provides a behavioral measure to assess emotions, cognitive processes, and social interaction [1]. Examples of applications of AFER include robotics, human–computer interface, behavioral science, animations and computer games, educational software, emotion processing, and fatigue detection. Due to multiple limitations and difficulties such as occlusion, lighting conditions, and variation of expressions across the population, or even for an individual, having an automatic system helps in creating intelligent visual media for understanding different expressions. Moreover, this understanding helps in building meaningful and responsive HCI interfaces.

Each AFER implements three main functions: face detection and tracking, feature extraction, and expression classification. The first attempt towards the AFER was

taken in 1978 by Suwa et al. [2] who presented a system for facial expression analysis from video, and tracking 20 points as features. Before that, only two ways existed for FER [3]: (a) human observer-based coding system which is subjective, time-consuming, and hard to standardize, and (b) electromyography-based systems which is invasive (needs sensors on the face). The muscle actions result in various facial behaviors and motions, and later on can be used to represent the corresponding facial expressions. These assumptions became the basis for developing the following systems for coding multiple facial expressions and emotions:

- 1 The Facial Action Coding System (FACS)—Ekman and Friesen [4].
- 2 The Facial Animation parameters (FAPs)—MPEG-4 standard, SNHC [5].

In the study of Ekman and Friesen [6], it was shown that those six emotions—anger, disgust, fear, happiness, sadness, and surprise—are “discriminable within any one literate culture”. Sometimes, a neutral expression is considered as a seventh expression. The FACS describes facial expressions in terms of action units (AUs). It

* Correspondence: apoursab@ucalgary.ca

¹Department of Electrical and Computer Engineering, University of Calgary, Canada ENA221, ICT Building, 2500 University Drive NW, Calgary T2N 1N4, Canada

Full list of author information is available at the end of the article

explains how to identify different facial expressions based on the application of varying facial muscles individually or in groups. It contains 46 AUs, which are basic facial movements, corresponding to different muscle activities. It is not an easy task to recognize AUs automatically, given an image or a video. There are two main approaches to AU recognition:

1. Processing 2D static images.
2. Processing image sequences.

The first one, which is more difficult than image sequence since less information is available, often uses feature-based methods. Using only one image for expression recognition needs robust and highly distinctive features to cope with variations in human subjects or imaging conditions [3]. There are several methods to process still images. One of them is PCA-based holistic representations and feed forward neural networks (NN) for classification proposed by Cottrell and Metcalfe [7]. Chen and Huang [8] used a clustering-based feature extraction to recognize only three facial expressions. Eigenface feature extraction accompanied by principal component analysis (PCA) is proposed by Turk and Pentland [9]. Holistic representations and NNs are applied to pyramid-structured images by Rahardja et al. [10]. Feng et al. [11] applied local binary pattern for feature extraction, and used a linear programming technique as the classifier. Deformable models were utilized by Lanitis et al. [12] to capture variations in shape and grey-level appearance. In the second approach, an image sequence displays one expression. The neutral face is used as a baseline face, and FER is based on the difference between the baseline and the following input face image. Preliminary work on facial expressions, by tracking the motion of 20 identified spots, has been done by Suma et al. [13]. Motion tracking of facial features in image sequences is performed by optical flow, and expressions are classified into six basic classes [14]. The Fourier transform was utilized for feature extraction, and a fuzzy C-means clustering was applied to build a spatiotemporal model for each expression in [15].

Facial coding is normally performed in two different ways: holistic and analytic. In the holistic approach, the face is treated as a whole. Different methods are presented in this approach including [16,17]: optical flow, Fisher linear discriminates, NN, active appearance models (AAMs), and Gabor filters. In the analytic approach, local features are used instead of the whole face, namely, fiducial points describe the position of important points on the face (e.g., eyes, eyebrows, mouth, nose, etc.), together with the geometry or texture features around these points [18].

Gabor filters are widely used in texture analysis. These filters model simple cells in the primary visual cortex. Zafeiriou and Pitas [19] showed the best performance of Gabor filters in both analytic and holistic approaches. Gabor filters have been used for expression classification in [4,20]. Although the Gabor filters show high performance in FER, the main problems using this filter is how to select the optimum one, in terms of scale and orientation. For example in [20], 40 filters (5 scales and 8 orientations) are used. Because of the large number of convolution operations, it needs large amounts of memory and computational cost. Moreover, with the small training samples, the dimensionality is really high [21]. Normally, two types of facial features are used: permanent and transient. Permanent features include eyes, lips, brows and cheeks, and transient features include facial lines, wrinkles, and furrows. The eyebrows and mouth play the main role in facial expressions. Pardas and Bonafonte [22] showed that expressions such as surprise, joy, and disgust have much higher recognition rate, since clear motion of the mouth and the eyebrows are involved.

In this article, both the combined texture and the geometric information of face fiducial points are used to code different expressions. Gauss–Laguerre (GL) wavelets are used for texture analysis and the positions of 18 fiducial points represent the deformation of the eyes, eyebrows, and mouth. The combination of these features is used for expression classification. The K-nearest neighbor (KNN) is used for classifying expressions based on closest training examples in the feature space. The rest of the article is organized as follows: in “G–L wavelets” section, a mathematical description of GL circular harmonic wavelets (CHW) is presented; feature extraction approach in addition to the classification method are mentioned in “The proposed approach” section; experimental results, using the JAFFE, the Cohn-Kanade, and the MMI face databases are reported in “Experiment results” section; finally, a conclusion is drawn in “Conclusion” section.

GL wavelets

The CHWs are polar-separable wavelets, with harmonic angular shape. They are steerable in any desired direction by simple multiplication with a complex steering factor, thus they are referred to as self-steerable wavelets. The CHWs were first introduced in [23] and utilize the concepts from circular harmonic functions (CHF) employed in optical correlations for rotation-invariant pattern recognition. A CHF is represented in polar coordinates [21] as

$$f_k^n(r, \theta) = V_k^n(r) e^{in\theta} \quad (1)$$

where n is the order, k is the degree of the CHF, and $V_k^n(\cdot)$ is the radial profile.

The same functions also appear in harmonic tomographic decomposition, and have been considered for the analysis of local image symmetry. CHFs have been employed for defining of rotation-invariant pattern signatures. A family of orthogonal CHWs, forming a multi-resolution pyramid referred to as the circular harmonic pyramid (CHP), is utilized for coefficient generation and coding. Each CHW, pertaining to the pyramid, represents the image by translated, dilated, and rotated versions of a CHF. At the same time, for a fixed resolution, the CHP orthogonal system provides a local representation of the given image around a point in terms of CHFs. The self-steerability of each component of the CHP can be exploited for pattern analysis in the presence of rotation (other than translation and dilation), in particular, for pattern recognition, irrespective of orientation.

CHFs, these are complex, polar separable filters, characterized by harmonic angular shape, which allows building rotationally invariant descriptors. A scale parameter is also introduced to perform a multi-resolution analysis. The GL filters from the family of orthogonal functions, satisfying the wavelet admissibility condition required for multi-resolution wavelet pyramid analysis, are used. Similar to Gabor wavelets, any image may be represented by translated, dilated, and rotated replicas of the GL wavelet. For a fixed resolution, the GL CHFs provide a local representation of the image in the polar coordinate system centered at a given point, named the pivot point. This representation is called the GL transform [24]. They are characterized by a CHF, which is a complex polar separable filter with a harmonic angular shape, represented in polar coordinates.

For a given image $I(x, y) \in L^2(\mathbb{R}^2, dx^2)$, the expression $I_p(r, \theta) = I(\tilde{x} + r \sin \theta, \tilde{y} + r \cos \theta)$ is the representation in the polar coordinate space centered at the pivot $\tilde{x}(\tilde{x}, \tilde{y})$. $I_p(\cdot)$ can be decomposed in terms of CHF, based on its periodic characteristic with respect to θ :

$$I_p(r, \theta) = \sum_n \sum_k V_k^n(r) e^{jn\theta} \quad (2)$$

where radial profile $V_k^n(\rho)$ is given by the Fourier integral:

$$V_k^n(r) = \frac{1}{2\pi} \int_0^{2\pi} I_p(r, \theta) e^{-jn\theta} d\theta \quad (3)$$

The expansion of radial profiles is represented by the series of weighted orthogonal functions which are GL CHFs:

$$\mathfrak{G}_k^n(r, \theta) = (-1)^K 2^{\frac{(n+1)}{2}} \pi^{\frac{|n|}{2}} \left[\frac{K!}{(|n| + K)!} \right]^{\frac{1}{2}} r^{|n|} L_K^n(2\pi r^2) e^{-\pi r^2} e^{jn\theta} \quad (4)$$

where $L_K^n(r)$ is the generalized Laguerre polynomial defined by

$$L_K^n(r) = \sum_{h=0}^K (-1)^K \binom{n+K}{K-h} \frac{r^h}{h!} \quad (5)$$

As any CHF, GL functions are self-steering, i.e., they are rotated by angle ϕ when multiplied by factor $e^{jn\phi}$. In particular, the real and imaginary parts of each GL function form a geometrical pair in phase quadrature. Moreover, GL functions are isomorphic to their Fourier transform. It is shown in [24] that each GL function defines an admissible dyadic wavelet. Thus, the redundant set of wavelets; corresponding to different GL functions, represent a self-steering pyramid, utilized for local and multiscale image analysis. The real part of the GL function is depicted in Figure 1a. An important feature, applicable to facial expression recognition, is that GL function with various degrees of freedom can be tuned to significant visual features. For example, for $n = 1$, GLs are tuned to edges, for $n = 2$ to ridges, for $n = 3$ to equiangular forks, for $n = 4$ to orthogonal crosses, irrespective of their actual orientation [25]. Given an image $I(x, y)$, for every site of the plane, it is possible to perform the GL analysis by convolving it with each properly scaled GL function as follows:

$$g_{jnK}(x, y) = \frac{1}{a2^j} L_K^n \left(\frac{r \cos \theta}{a2^j}, \frac{r \sin \theta}{a2^j} \right) \quad (6)$$

where $a2^j$ are the dyadic scale factors. Figure 1b shows a plot of GL-CHFs for a fixed dyadic scale factor and variables n and k .

The proposed approach

In this section, the algorithmic steps of proposed approach are explained. For each input image, the face area is localized first. Then, the features are extracted based on GL filters, and, finally, the KNN classification is used for expression recognition.

Preprocessing

Preprocessing is normally performed before feature extraction for the FER, in order to increase system performance. The aim of this step which includes scaling, intensity normalization, and size equalization, is to have images, which only contain a face, expressing a certain emotion. Sometimes, histogram equalization is also used to adjust image brightness and contrast. To normalize the face, the image with the neutral expression is scaled, so that it has a

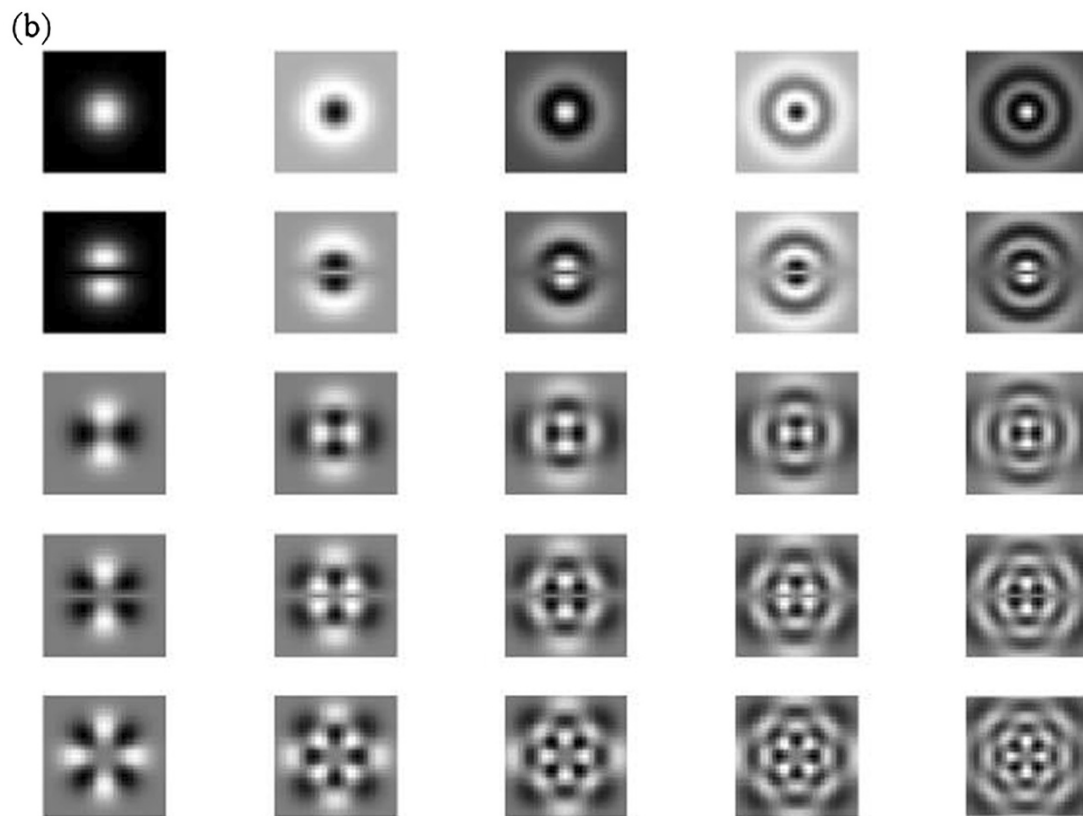
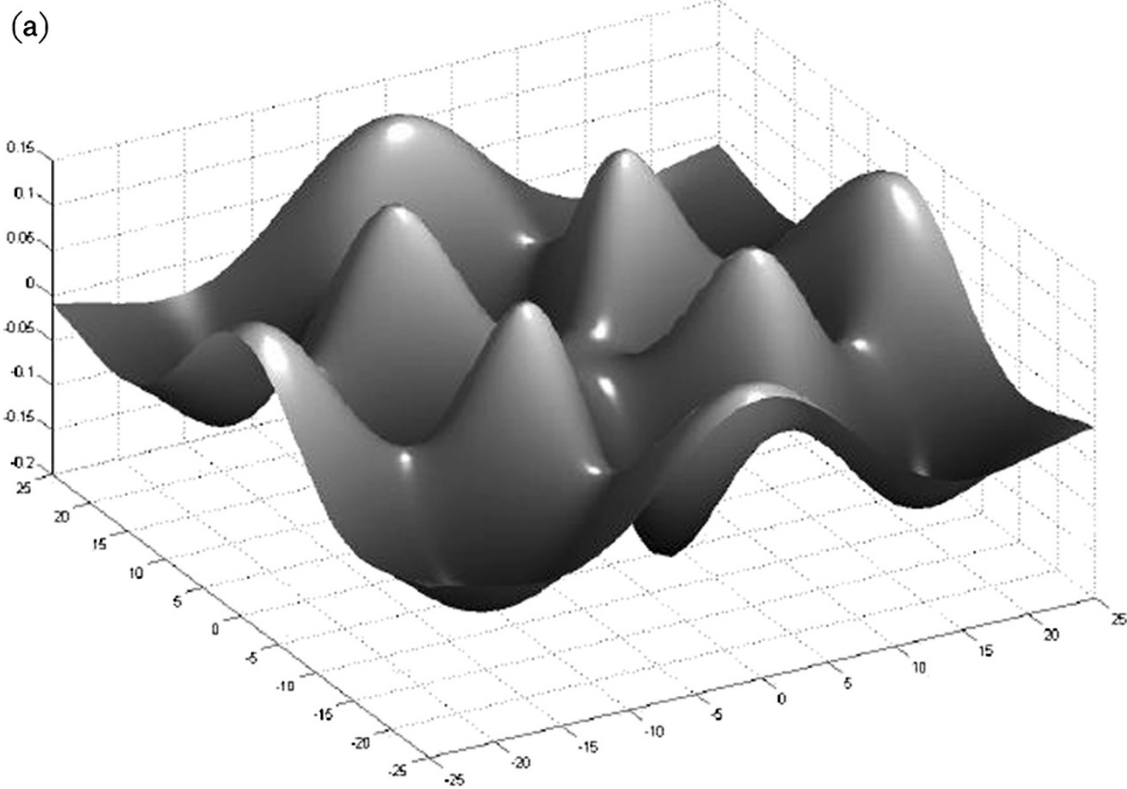


Figure 1 (a) Real part of GL function; $n = 4, K = 1, j = 2$. (b) Real part of GL CHF. Variation of filter in spatial domain with fixed scale, $k = 0, 1, \dots, 4$, and $n = 1, 2, \dots, 5$.

fixed distance between the eyes. No intensity normalization has been considered, since the GL filters can extract an abundance of features without any preprocessing.

For face localization, we used the well-known algorithm by Viola-Jones. This method is based on the Haar-like features and the AdaBoost learning algorithm. The localized face image is cropped automatically and resized to 128×96 . The next step is to normalize the geometry, so that the recognition method is robust to individuals' differences. For the normalization, we need to extract the eye locations. Figure 2 shows an example of the normalization procedure. The output of this step is directly used for textural feature extraction. In this experiment, if the face was not cropped well, it was done manually. The purpose of this article is to propose a new feature extraction method for facial expression recognition.

To extract facial features, the AAM is utilized. It is widely used in face recognition and expression classification, due to its remarkable performance in extracting face shape and texture information. AAM [12] contains both a statistical model and texture information of the face, and performs matching via finding the model parameters. These minimize the difference between the image and the synthesized model. We used 18 fiducial points to model the face and distinguish facial expressions. The features to distinguish the latter are explained in section "AAM". In our experiment, the AAM model has been created using different images from three databases with different expressions. All images were roughly resized and cropped to 256×256 . After creating the AAM, the eye positions in each image is automatically extracted, and the line, which connects the inner corner of the eyes, is used for normalization.

AAM

AAM [12] is an algorithm for matching a statistical shape model to an image with both shape and appearance variations. For example in facial expression recognition, these deformations are both facial expression changes and pose variations along with the texture variations caused by illuminations. These variations are

represented by a linear model like PCA. So, the main purpose of the AAM is first to define a model and then finding the best matched parameters between the given new image and built model using a fitting algorithm.

Normally, the fitting algorithm is repeated until the parameters of shape and appearance satisfy particular values. The shape model is created by combining the vectors constructed from the points of the labeled images.

$$s = s_0 + \sum_{i=1}^m p_i s_i \quad (7)$$

where p_i is the parameters of shape. The mean shape s_0 and m shape basis vectors s_i are obtained by using PCA for training data. They are the m eigenvectors which correspond to the m largest eigenvalues. Before applying PCA, the landmark points are normalized. The appearance variation is represented by a linear combination of a mean appearance $A_0(x)$ and n appearance basis vectors $A_i(x)$ as

$$A(x) = A_0(x) + \sum_{i=1}^n \alpha_i A_i(x) \quad (8)$$

where α is the appearance parameter. After finding the shape and appearance parameters, a piecewise affine warp is used to construct the AAM by locating each pixel of appearance onto the inner side of the current shape. The goal is to minimize the difference between the warped image and the appearance image.

Feature extraction

The feature vector consists of two types: textural features, which are extracted globally by applying the GL filter, and the geometric information of local fiducial points.

Textural feature extraction

To extract facial features we used GL functions, which provide a self-steering pyramidal analysis structure. By the proper choice of the GL function parameters (scale,

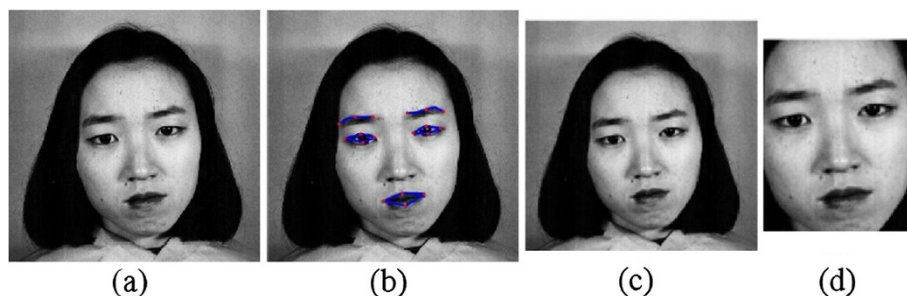


Figure 2 Normalization procedure (left to right): (a) input image (from the JAFFE database), (b) the extracted AAM fiducial points, (c) normalized image to have fixed distance between eyes, (d) the localized and resized face.

order, and degree of CHE, which is explained further), it is possible to generate a set of redundant wavelets, and, thus, an accurate extraction of complex texture features of facial expression. Redundancy of the wavelet transform and a higher degree of freedom in the selected parameters of the GL function makes it highly suitable for facial texture extraction, compared to those of the Gabor wavelets. The self-steering pyramid structure of GL-based analysis is superior to Gabor wavelets, due to its ability to choose parameters of GL functions. To take advantage of the degrees of freedom, provided by the GL function, parameters of filters have to be tuned to significant visual features and texture patterns, so that it can extract desirable frequency information of facial texture patterns. In our experiment, it was found that the best results are obtained for $n = 2$, via running several simulations, in which the filters are convolved directly with the face image. Other parameters (scale and degree) are also adjusted in the same manner. The best results have been obtained for $a = 2, k = 1$. The output of the filtered image has the same size as the input image 128×96 and is complex-valued. Figure 3 shows the examples of GL filtering with various parameters a and k . Unlike the

Gabor filter, there is no need to construct multiple filters, and a single tuned GL filter is sufficient for feature selection. The size of the textural feature vector ($128 \times 96 = 12288$) is quite large for fusion with geometric information points. If the dimension of the input vector is large, and the data are highly correlated, there are several methods to remove redundancy, such as PCA. During PCA, the components of input vectors are orthogonalized, which means they are no more correlated with each other. The components are put in order, so that ones with largest variation come first and those with low variation are eliminated. The data are usually normalized before performing PCA to have zero mean and unity variance. In our case, the size of the feature vector, after down-sampling, is 3072 (by factor 4), which is further reduced to 384 samples per image using the PCA.

Geometric feature extraction

As mentioned in section “Preprocessing”, 18 fiducial points are put together to construct the model. These points are extracted automatically, based on the AAM model. The coordinates of these fiducial points are used to calculate 15 Euclidean distances. Different expressions result in different

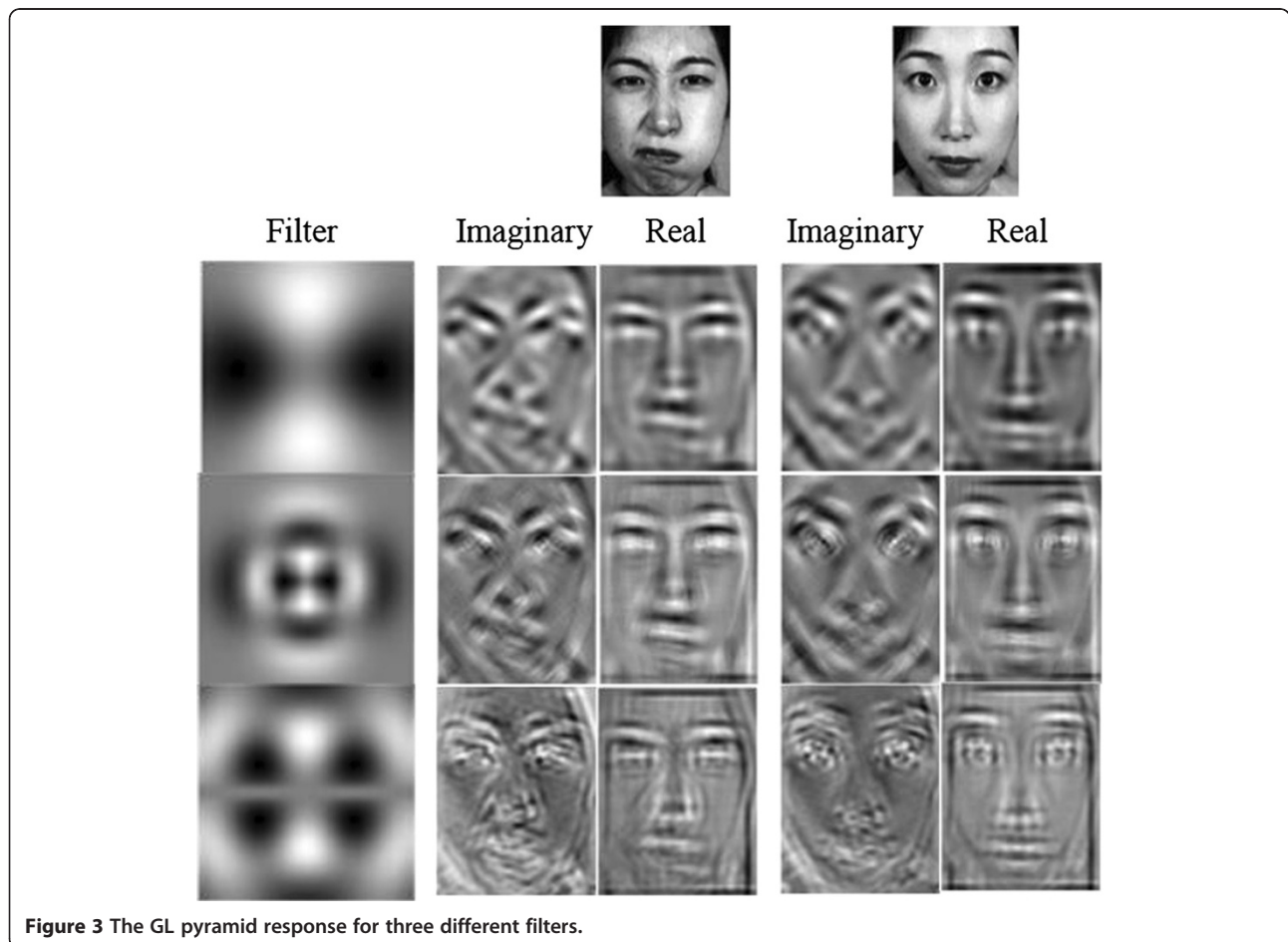
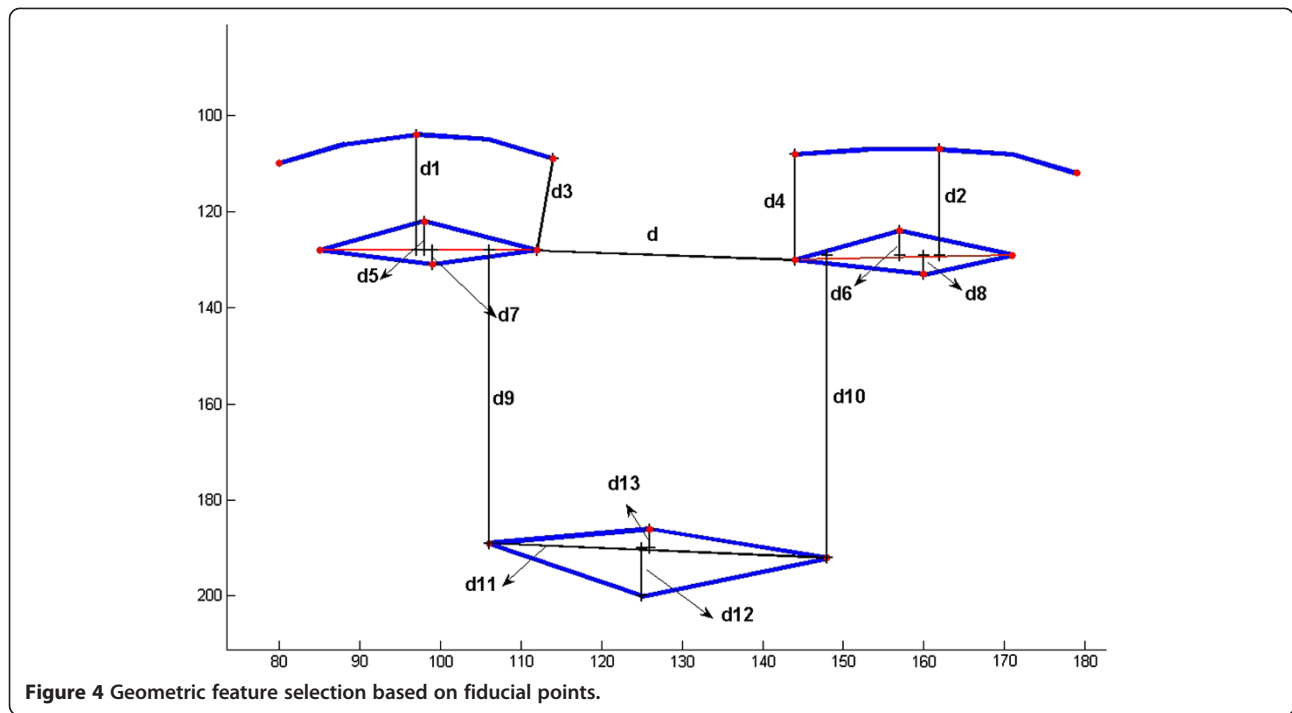


Figure 3 The GL pyramid response for three different filters.



deformations of the corresponding facial components, especially near the eyes and mouth. The selected geometric feature extractions are performed as follows.

1. The AAM is applied to extract the 18 points. The distances are labeled by d 's as shown in Figure 4.
2. For the upper portion of the face, ten distances are calculated, according to Table 1.
3. For the lower portion of the face, five distances are calculated, according to Table 1.

The final feature vector is a combination of two features. Since the dimension of the texture feature vector is 384,

and the dimension of the geometric feature vector is 15, the total size is 399. During the simulations, it was observed that geometric features are more important than texture ones. To find the appropriate weight coefficients for both types of features, the average recognition rate versus different weight coefficient for geometrical features have been monitored. Let us assume that the textural feature is F_T and the geometrical feature is F_G . The final feature vector is $F = \alpha F_T + \beta F_G$. The average recognition rate has been obtained using the "Leave-One-Out" approach in three trials (see section "Experiment results"). In each trial, and for each database, one random image is used for testing and the rest utilized for training. In this case, the three

Table 1 Upper (distances 1–10) and lower (distances 11–15) face geometric distance

Meaning	Distance	Meaning	Distance
left inner brow-left inner eye corner	$P_1 = \frac{d_3E - d_3N}{d_3N}$	right eye height	$P_6 = \frac{(d_6E + d_5E) - (d_6N + d_5N)}{(d_6N + d_5N)}$
right inner brow-right inner eye corner	$P_2 = \frac{d_4E - d_4N}{d_4N}$	left top eye point-line connecting left eye corners	$P_7 = \frac{d_5E - d_5N}{d_5N}$
left top brow-line connecting left eye corners	$P_3 = \frac{d_1E - d_1N}{d_1N}$	right top eye point-line connecting right eye corners	$P_8 = \frac{d_6E - d_6N}{d_6N}$
right top brow-line connecting right eye corners	$P_4 = \frac{d_2E - d_2N}{d_2N}$	left bottom eye point-line connecting left eye corners	$P_9 = \frac{d_7E - d_7N}{d_7N}$
left eye height	$P_5 = \frac{(d_5E + d_7E) - (d_5N + d_7N)}{(d_5N + d_7N)}$	right bottom eye point-line connecting right eye corners	$P_{10} = \frac{d_8E - d_8N}{d_8N}$
Mouth height	$P_{11} = \frac{(d_{12}E + d_{13}E) - (d_{12}N + d_{13}N)}{(d_{12}N + d_{13}N)}$	Mouth width	$P_{12} = \frac{d_{11}E - d_{11}N}{d_{11}N}$
left lip corner-line connecting left eye corners	$P_{13} = \frac{d_9E - d_9N}{d_9N}$	right lip corner-line connecting right eye corners	$P_{14} = \frac{d_{10}E - d_{10}N}{d_{10}N}$
top lip-line connecting lip corners	$P_{15} = \frac{d_{13}E - d_{13}N}{d_{13}N}$		

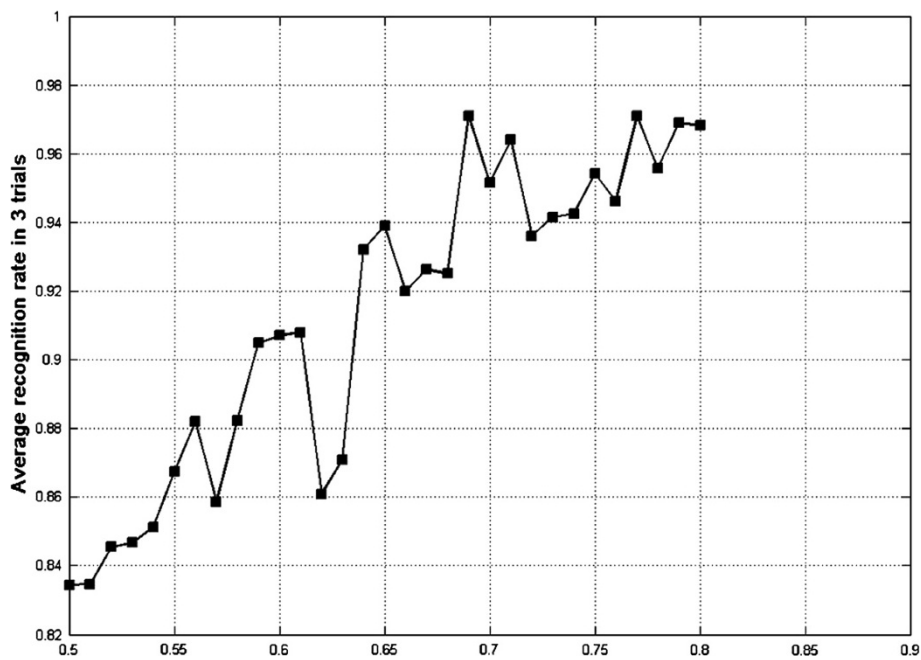


Figure 5 Variation of coefficient for geometric features versus average recognition rate.

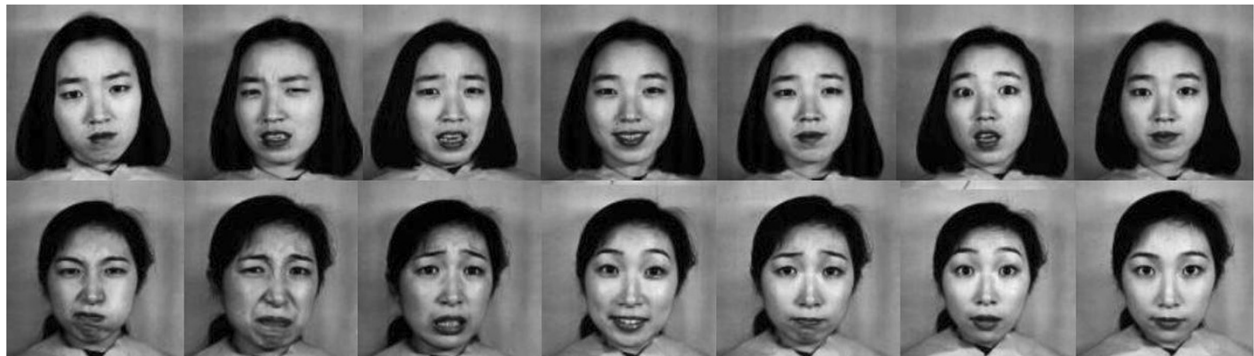


Figure 6 Various expressions from the JAFFE database (left to right): anger, disgust, fear, happiness, sadness, surprise, and neutral.



Figure 7 Examples of various expressions from the Cohn-Kanade database.



Figure 8 The sample face expression images from the MMI database.

Table 2 Recognition accuracy (%) on the JAFFE database for different approaches

Expression	Leave-One-Out	Cross validation	Expresser-based segmentation
Anger	94.3	92.8	89.3
Disgust	95.7	94.6	90.7
Fear	96.0	95.0	91.1
Happiness	98.2	96.9	92.6
Sad	96.7	94.2	90.2
Surprise	98.2	96.3	92.3
Neutral	97.9	95.5	91.7
Average	96.71	95.04	91.12

trials, generally speaking, have different sets for train/test schemes. The coefficient for geometric features varied from 0.5 by step-size 0.01. Figure 5 shows the average recognition rate versus geometric coefficient weight. The best average recognition rate, within coefficients, was $\beta = 0.69$ and $\alpha = 0.31$.

Classification

KNN is a well-known instance-based classification algorithm [26], which does not make any assumptions on the underlying data distribution. The similarity between the test sample and the other samples, used in training, is calculated, and k most similar set samples are determined. The class of the test sample is then found, based on the classes of its KNNs.

This classification suits the multi-class classification, in which the decision is based on a small neighborhood of similar objects. In the classification procedure, the training data are first plotted in n -dimensional space, where n is the number of features. Each of these consists of a set of vectors labeled with their associated class (arbitrary number of classes). The number k defines how many neighbors influence the classification. Based on the suggestion made in [26], the better classification is obtained when $k = 3$. This suggestion was based on different experiments and observing the classification rate on JAFFE database. The same classifier is used for the Cohn-Kanade and the MMI database as well.

Table 3 Confusion matrix for the Leave-One-Out method (the JAFFE database)

	AN	FE	SU	DI	HA	SA	NE	Total
AN	94.3%	0	0.7	3.2%	0	1.8%	0	30
FE	1%	96%	2%	1%	0	0	0	32
SU	0	0.5%	98.2%	0	1.3%	0	0	30
DI	1.1	0	0.5%	95.7%	0	2.7%	0	29
HA	0	0	0	0	98.2%	0	1.8%	31
SA	2%	0	0.3%	1%	0	96.7%	0	31
NE	0	0	0	0	0	2.1%	97.9%	30

Table 4 Comparison with other methods on the JAFFE database

Method	Recognition rate (average%)
Lyons et al. [31]	92.00
Zhi and Ruan [33]	95.91
Zhang et al. [34]	90.34
Liejun et al. [35]	95.7
Shih et al. [36]	95.71
Zhao et al. [37]	93.72
Guo and Dyer [38]	91.00
Proposed	96.71

Table 5 Recognition accuracy (%) on the Cohn-Kanade database for different approaches

Expression	Leave-One-Out	Cross validation	Expresser-based segmentation
Anger	88.25	87.03	82.35
Disgust	94.9	91.58	88.65
Fear	92.28	90.98	87.12
Happiness	97.82	96.92	91.05
Sad	87.87	84.58	81.11
Surprise	92.08	91.23	86.32
Average	92.2	90.37	86.1

Experiment results

To evaluate the performance of the proposed method, the JAFFE image database [27], the Cohn-Kanade, and the MMI databases have been used. Eighteen fiducial points have been obtained via the AAM model, and two types of information have been extracted: geometric and textural. MATLAB was used for implementation.

The JAFFE database contains 213 images with a resolution of 256×56 pixels. Six basic expressions, in addition to the neutral face (seven in total), are considered in this database. These expressions are happiness, sadness, surprise, anger, disgust, and fear. The images were taken from 10 Japanese female models, and each emotion was subjectively tested on 60 Japanese volunteers. There are three samples, corresponding to each facial expression of each person. Figure 6 shows examples of each expression. In the JAFFE database, each subject posed 3 or 4 times for each expression.

The Cohn-Kanade database [28], which is widely used in literature for facial expression analysis, consists of approximately 500 image sequences from posers. Each sequence goes from neutral to target display, with the last frame being AU coded. The subjects range in age from 18 to 30 years where 65% are females, 15% are African-American, and 3% are Asian or Latino. The images contain six different facial expressions: anger, disgust, fear, happiness, sadness, and anger. Figure 7 shows examples of some expressions.

The MMI Facial Expression Database [29] was created by the Man-machine Interaction Group, Delft University

of Technology, Netherlands. This database is initially established for research on machine analysis of facial expressions. The database consists of over 2,900 videos and high-resolution still images of 75 subjects of both genders, who range in age from 19 to 62 years and have either a European, Asian, or South American ethnic background. These samples show both non-occluded and partially occluded faces, with or without facial hair and glasses. In our experiments, 96 image sequences were selected from the MMI database. The only selection criterion is that a sequence can be labeled as one of the six basic emotions [30]. The sequences come from 20 subjects, with 1–6 emotions per subject. The neutral face and three peak frames of each sequence (hence, 384 images in total) were used for 6-class expression recognition. Some sample images from the MMI database are shown in Figure 8.

Three different methods were selected to verify the accuracy of this system:

1. *“Leave-One-Out” cross-validation*: For each expression from each subject, one image is left out, and the rest are used for training [26].
2. *Cross-validation*: the database is randomly partitioned to ten distinct segments, and nine partitions are used for training, with the remaining partition used to test performance. The procedure is repeated so that every equal-sized set is used once as the test set. Finally, an average of ten experiments is been reported [31].
3. *Expresser-based segmentation*: the database is divided into several segments; each of them corresponds to a subject. For the JAFFE database, 213 expression images, posed by 10 subjects, are partitioned into 10 segments, each corresponding to one subject [32]. For the Cohn-Kanade database, 375 video sequences are been used, that is, over 4,000 images. Nine out of ten segments are used for training and the tenth for testing. It is repeated, so each of the ten segments is used in testing. The average results for those ten experiments are been reported.

Table 6 Confusion matrix for the Leave-One-Out method (the Cohn-Kanade database)

	AN (%)	FE (%)	SU (%)	DI (%)	HA (%)	SA (%)
AN	88.25	1.21	0.89	1.08	0.00	8.57
FE	0.87	92.28	1.98	2.17	2.21	0.49
SU	1.87	1.43	92.08	1.35	1.31	1.96
DI	1.65	0.67	0.00	94.9	0.00	2.78
HA	0.00	0.00	2.18	0.00	97.82	0.00
SA	6.54	4.63	0.00	0.96%	0.00	87.87

Table 7 Comparison of facial expression recognition for the Cohn-Kanade database

Method	Number of selected video sequences	Recognition rate (average%)
Zhan et al. [39]	300	90.4
Shan et al. [40]	320	92.1
Bartlett et al. [41]	313	86.9
Littlewort et al. [42]	313	93.8
Yang et al. [43]	352	92.23
Tian [44]	375	93.8
Aleksic and Katsaggelos [45]	284	93.66
Zafeiriou and Pitas [19]	374	97.1
Kotsia and Pitas [46]	–	99.7%
Proposed	374	92.20

JAFFE database

Table 2 shows the average success rate for different approaches. The confusion matrix for the “Leave-One-Out” method is presented in Table 3. For the average recognition rate, nine out of ten expression image classes are used for training with the last one being the testing set each time. This procedure is repeated for each subject.

The confusion matrix is a 7×7 matrix containing information of actual class label in both rows and columns. The diagonal entries are the rounded average successful recognition rates in ten trials, while the off-diagonal entries correspond to misclassifications. The total recognition rate is 96.71%; the best rate is for surprise and happiness expressions, and the lowest one is for anger. The performance of the proposed method has been compared against some published methods in Table 4.

Cohn-Kanade database

Table 5 shows the average success rate for different approaches. The confusion matrix for the “Leave-One-Out” method is also presented in Table 6. Different number of images has been selected for experiments in literature, and images are selected based on the different criteria (see Table 7). In this experiment, 375 image sequences have been selected from 97 subjects so that

the criterion was to be that of a sequence labeled as one of the six basic emotions, with the video clip being longer than ten frames. The total recognition rate is 92.2%; the best rate is for the happiness expression, and the lowest one for sadness. The performance of the proposed method has been compared against some published methods in Table 7. Although several frames from each video sequence are used, we consider them as “static” images without using any temporal information.

MMI database

Table 8 shows the average success rate for different approaches. The total recognition rate is 87.6%; the best rate is for the happiness expression, and the lowest one being for sadness.

The experimental results show that the proposed method meets the criteria of accuracy and efficiency for facial expression classification. It outperforms, in terms of accuracy, some other existing approaches that used the same database. The average recognition rate of the proposed approach is 96.71%, when using “Leave-One-Out” method, and 95.04% when using cross-validation for estimating its accuracy on the JAFFE database. For the Cohn-Kanade database, the average recognition rate of the proposed approach is 92.20%, when using “Leave-One-Out” method, and 90.37% when using the cross-validation for estimating its accuracy. For the MMI

Table 8 Recognition accuracy (%) on the MMI database for different approaches

Expression	Leave-One-Out	Cross validation	Expresser-based segmentation
Anger	86.14	85.08	80.11
Disgust	85.22	86.4	78.22
Fear	89.91	84.42	81.32
Happiness	91.1	88.81	83.21
Sad	83.44	81.79	77.11
Surprise	90.2	89.34	81.02
Average	87.66	85.97	80.16

database, the average recognition rate of the proposed approach is 87.66%, when using the “Leave-One-Out” method, and 85.97% when using cross-validation for estimating its accuracy. Few articles reported the accuracy on emotion recognition on the MMI. Most of them reported the recognition rate on the AU. Sánchez et al. [47] achieved 92.42% but it is not clear how many video sequences were used. Cerezo et al. [48] reported 92.9% average recognition rate on 1,500 still images of mixed MMI and CK databases. Shan et al. [30] used 384 images from the MMI, and the average recognition rate of 86.9% was reported.

For the “Leave-One-Out” procedure in Table 5, all image sequences are divided into six classes, each corresponding to one of the six expressions. Four sets, each containing 20% of the data for each class, chosen randomly, were created to be used as training sets, while the other 20% were used as the test set.

The procedure of classification is repeated five times. In each cycle, the samples in the testing set are included into the current training set. The new set of samples (20% of the samples for each class) is again formed to have a new test set, and the remaining ones are the new training set. Finally, the average classification rate is the mean of the success rate in classification.

Conclusion

This article proposes a combined texture/geometric feature selection for facial expression recognition. The GL circular harmonic filter is applied, for the first time, to facial expression identification. The advantage of this filter is its rich frequency extraction capability for texture analysis, as well as being a rotation-invariant and a multiscale approach. The geometric information of fiducial points is added to the texture information to construct the feature vector. Given a still expression image, normalization is performed first. The extracted features are passed through a KNN classifier. Experiments showed that the selected features represent the facial expression effectively, demonstrating an average success rate of 96.71, 92.2, and 87.66% when following the “Leave-One-Out” strategy for accuracy estimation, as well as 95.04, 90.37, and 85.97% when following the cross-validation method. These are comparable with the results, reported for other approaches on both databases, namely, the presented results demonstrate better success rate for the JAFFE database, and have the same success range as the approaches for the Cohn-Kanade database. Further development of the proposed approach includes perfecting the local and global feature selections, as well as testing using other classification techniques.

Author details

¹Department of Electrical and Computer Engineering, University of Calgary, Canada ENA221, ICT Building, 2500 University Drive NW, Calgary T2N 1N4, Canada. ²Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada.

Received: 1 November 2011 Accepted: 3 September 2012

Published: 25 September 2012

References

1. M Yuki, WW Maddux, T Masuda, Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *J. Exp. Soc. Psychol.* **43**(2), 303–311 (2007)
2. M Suwa, N Sugie, K Fujimora, *A preliminary note on pattern recognition of human emotional expression, in Proceedings of the Fourth International Joint Conference on Pattern Recognition* (, Kyoto, Japan, 1978), pp. 408–410
3. S Bashyal, GK Venayagamoorthy, Recognition of facial expressions using Gabor wavelets and learning vector quantization. *Eng. Appl. Artif. Intell.* **21**, 1056–1064 (2008)
4. P Ekman, *WV Friesen, Manual for the Facial Action Coding System* (Consulting Psychologists Press, Palo Alto, CA, 1977)
5. Audio (MPEG Mtg, Atlantic City, 1998), MPEG Video and SNHC, *Text of ISO/IEC FDIS 14 496-3*. (). Doc. ISO/MPEG N2503
6. P Ekman, WV Friesen, Constants across cultures in the face and emotions. *J. Personal Soc. Psychol.* **17**(2), 124–129 (1971)
7. G Cottrell, J Metcalfe, Face, Gender and Emotion Recognition Using Holons, in *Advances in Neural Information Processing Systems*, ed. by K. Morgan, M. San, 3rd edn. (, 1991), pp. 564–571
8. X Chen, T Huang, Facial expression recognition: a clustering based approach. *Pattern Recognit. Lett.* **24**, 1295–1302 (2003)
9. M Turk, A Pentland, Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**, 71–86 (1991)
10. A Rahardja, A Sowmya, W Wilson, A neural network approach to component versus holistic recognition of facial expressions in images. *Intell. Robots Comput. Vis. X: Algorithms and Techniques* **1607**, 62–70 (1991)
11. X Feng, M Pietikäinen, A Hadid, Facial expression recognition based on local binary patterns. *Pattern Recognit. Image Anal.* **17**(4), 592–598 (2007)
12. A Lanitis, C Taylor, T Cootes, Automatic interpretation and coding of face images using flexible models. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 743–756 (1997)
13. M Suma, N Sugie, K Fujimora, *A preliminary note on pattern recognition of human emotional expression, in Proceedings of the 4th International Joint Conference on Pattern Recognition* (, Kyoto, Japan, 1978), pp. 408–410
14. Y Yacoob, L Davis, *Recognizing faces showing expressions, in International Workshop Automatic Face and Gesture Recognition* (, Zurich, Switzerland, 1995), pp. 278–283
15. T Xiang, MKH Leung, SY Cho, Expression recognition using fuzzy spatio-temporal modeling. *Pattern Recognit.* **41**(1), 204–216 (2008)
16. I Essa, A Pentland, Coding, analysis, interpretation and recognition of facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 757–763 (1997)
17. IR Fasel, MS Bartlett, JRA Movellan, *A comparison of Gabor filter methods for automatic detection of facial landmarks* (IEEE 5th International Conference on Automatic Face and Gesture Recognition, Washington, DC, 2002), pp. 242–248
18. B Fasel, J Luetttinb, Automatic facial expression analysis: a survey. *Pattern Recognit.* **36**(1), 259–275 (2003)
19. S Zafeiriou, I Pitas, Discriminant graph structures for facial expression recognition. *IEEE Trans. Multimed.* **10**(8), 1528–1540 (2008)
20. CC Lee, CY Shih, *Gabor feature selection for facial expression recognition* (International Conference on Signals and Electronic Systems, Gliwice, Poland, 2010), pp. 139–142
21. H Deng, J Zhu, MR Lyu, I King, *Two-stage multi-class AdaBoost for facial expression recognition* (Proceedings of IJCNN07, Orlando, USA, 2007), pp. 3005–3010
22. M Pardas, A Bonafonte, Facial animation parameters extraction and expression recognition using Hidden Markov Models. *Signal Process: Image Commun* **17**, 675–688 (2002)
23. G Jacovitti, A Neri, Multiscale image features analysis with circular harmonic wavelets. *Proc. SPIE: Wavelets Appl. Signal Image Process.* **2569**, 363–372 (1995)
24. L Capdiffero, V Casieri, A Laurenti, G Jacovitti, Multiple feature based multiscale image enhancement. Greece, in *Digital Signal Processing Conference* **2**, 931–934 (2002)

25. H Ahmadi, A Poursaberi, A Azzadeh, M Kamarei, An efficient iris coding based on Gauss-Laguerre wavelets, in *2nd IAPR/IEEE International Conference on Biometrics*, Seoul, South Korea **4642**, 917–926 (2007)
26. A Sohail, P Bhattacharya, *Classification of facial expressions using k-nearest neighbor classifier*, 4418th edn. (Computer Vision Computer Graphics Collaboration Techniques, 2007), pp. 555–566
27. M Lyons, S Akamatsu, M Kamachi, J Gyoba, *Coding facial expressions with Gabor wavelets* (Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998), pp. 200–206
28. T Kanade, JF Cohn, Y Tian, *Comprehensive database for facial expression analysis* (Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000), pp. 46–53
29. M Pantic, MF Valstar, R Rademaker, L Maat, *Web-based database for facial expression analysis* (Proceedings of the IEEE International Conference on Multimedia and Expo (ICME'05), Amsterdam, Netherlands, 2005), pp. 317–321
30. C Shan, G Shaogang, PW McOwan, Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis. Comput.* **27**, 803–816 (2009)
31. M Lyons, J Budynek, S Akamatsu, Automatic classification of single facial images. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 1357–1362 (1999)
32. X Feng, B Lv, Z Li, J Zhang, *A novel feature extraction method for facial expression recognition* (Proceeding of JCS, Taiwan, 2006), pp. 371–375
33. R Zhi, Q Ruan, Facial expression recognition based on two dimensional discriminant locality preserving projections. *Neuro Comput.* **71**, 1730–1734 (2008)
34. Z Zhang, M Lyons, M Schuster, S Akamatsu, *Comparison between geometry based and Gabor wavelet based facial expression recognition using multi layer perceptron* (Proceeding 3rd International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998), pp. 454–459
35. W Liejun, Q Xizhong, Z Taiyi, Facial expression recognition using improved support vector machine by modifying kernels. *Inf. Technol. J.* **8**(4), 595–599 (2009)
36. FY Shih, C Chuang, PSP Wang, Performance comparisons of facial expression recognition in Jaffe database. *IJPRAI*, 445–459 (2008)
37. L Zhao, G Zhuang, X Xu, *Facial expression recognition based on PCA and NMF* (Proceeding of the 7th World Congress on Intelligent Control and Automation, Chongqing, China, 2008), pp. 6822–6825
38. G Guo, CR Dyer, Learning from examples in the small sample case: face expression recognition. *IEEE Trans. Syst. Man Cybern. B* **35**(3), 477–488 (2005)
39. Y Zhan, J Ye, D Niu, P Cao, Facial expression recognition based on Gabor wavelet transformation and elastic templates matching. *Int. J. Image Graph.* **6**(1), 125–138 (2006)
40. C Shan, S Gong, PW McOwan, Robust facial expression recognition using local binary patterns, in *Proceeding of ICIP05*, Genoa, Italy **2**, 370–373 (2005)
41. MS Bartlett, G Littlewort, I Fasel, JR Movellan, Real time face detection and facial expression recognition: development and applications to human computer interaction, in *IEEE Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin **5**, 53–53 (2003)
42. G Littlewort, M Bartlett, I Fasel, J Susskind, J Movellan, *Dynamics of facial expression extracted automatically from video*, 5th edn. (Proceeding of IEEE Conf. Computer Vision and Pattern Recognition, Workshop on Face Processing in Video, New York, USA, 2004), pp. 80–88
43. P Yang, Q Liu, DN Metaxas, *Exploring facial expressions with compositional features* (Proceeding of CVPR, San Francisco, USA, 2010), pp. 2638–2644
44. Y Tian, *Evaluation of face resolution for expression analysis* (Proceeding of IEEE Workshop Face Processing in Video, Washington, DC, USA, 2004), pp. 82–82
45. SP Aleksic, KA Katsaggelos, Automatic facial expression recognition using facial animation parameters and multi-stream HMMS. *IEEE Trans. Inf. Forensics Secur.* **1**(1), 3–11 (2006)
46. I Kotsia, I Pitas, Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Trans. Image Process* **16**(1), 172 (2007)
47. A Sánchez, JV Ruiz, M Ana, Belén, AS Montemayor, H Javier, P Juan José. Differential optical flow applied to automatic facial expression recognition. *Neurocomputing* **74**(8), 1272–1282 (2011)
48. E Cerezo, I Hupont, S Baldassarri, S Ballano, Emotional facial sensing and multimodal fusion in a continuous 2D affective space. *Ambient Intell. Hum. Comput.* **3**, 31–46 (2012)

doi:10.1186/1687-5281-2012-17

Cite this article as: Poursaberi et al.: Gauss–Laguerre wavelet textural feature fusion with geometrical information for facial expression identification. *EURASIP Journal on Image and Video Processing* 2012 **2012**:17.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
