Scientific Research Publishing

Journal of Signal and Information Processing, 2017, 8, 132-151
http://www.scirp.org/journal/jsip
ISSN Online: 2159-4481
ISSN Print: 2159-4465

Facial Expression Recognition Based on Local Fourier Coefficients and Facial Fourier Descriptors

Gibran Benitez-Garcia, Tomoaki Nakamura, Masahide Kaneko

Department of Mechanical Engineering and Intelligent Systems, The University of Electro-Communications, Tokyo, Japan
Email: gibran@radish.ee.uec.ac.jp

How to cite this paper: Benitez-Garcia, G., Nakamura, T. and Kaneko, M. (2017) Facial Expression Recognition Based on Local Fourier Coefficients and Facial Fourier Descriptors. *Journal of Signal and Information Processing*, **8**, 132-151.
https://doi.org/10.4236/jsip.2017.83009

Received: July 1, 2017
Accepted: July 28, 2017
Published: July 31, 2017

Abstract

The recent boom of mass media communication (such as social media and mobiles) has boosted more applications of automatic facial expression recognition (FER). Thus, human facial expressions have to be encoded and recognized through digital devices. However, this process has to be done under recurrent problems of image illumination changes and partial occlusions. Therefore, in this paper, we propose a fully automated FER system based on Local Fourier Coefficients and Facial Fourier Descriptors. The combined power of appearance and geometric features is used for describing the specific facial regions of eyes-eyebrows, nose and mouth. All based on the attributes of the Fourier Transform and Support Vector Machines. Hence, our proposal overcomes FER problems such as illumination changes, partial occlusion, image rotation, redundancy and dimensionality reduction. Several tests were performed in order to demonstrate the efficiency of our proposal, which were evaluated using three standard databases: CK+, MUG and TFEID. In addition, evaluation results showed that the average recognition rate of each database reaches higher performance than most of the state-of-the-art techniques surveyed in this paper.

Keywords

Facial Expression Recognition, Fourier Coefficients, Fourier Descriptors, Facial Region Segmentation, Partial Occlusion

1. Introduction

Facial expressions of emotions are defined by facial muscle movements which represent specific human emotions. Psychologists have established the facial expressions of emotions as six basic and universally recognized expressions: anger, disgust, fear, happiness, sadness and surprise [1].

On the other hand, ongoing researches of computer vision and machine learning try to find a suitable way to encode the facial representations which define human emotions. Thus, a complex Human-Computer Interaction (HCI) could be attained. Formally, automatic facial expression recognition (FER) is the field in charge of analyzing and recognizing facial feature changes from visual information (*i.e.* spatial or spatio-temporal representations). Some of the applications of automated and real-time FER systems include health-care, customer satisfaction analysis, virtual reality, smart environments, video-conferencing, human emotion analysis, cognitive science, and more [2].

FER systems can be categorized as spatial or spatio-temporal [3]. Spatial representations process static images, where information of only one frame is utilized for recognizing the shown expression. Whereas, spatio-temporal approaches consider a set of consecutive images for the recognition process, *i.e.* information contained in a sequence of frames. It is worth noting that the neutral face could be used as a baseline face for both categories. Another classification of FER systems can be based on terms of features, defined as appearance or geometric [4]. Appearance features represent the skin texture of the face and its changes (wrinkles and creases), meanwhile geometric features represent the shape of the face by using specific feature points from different facial parts. Some of the techniques which have been successfully applied to appearance-based feature extraction are Bag of Words [5], Gabor [6], LDA [7], LBP [8] and recently Convolutional Neural Networks (CNN) [9]. On the other hand, the methods applied for geometric features are AAM [10], EBGM [11], concatenation [12] and straight-line distances [13] of fiducial points. It is worth noting that survey papers ([3] and [4]) mention that approaches which combine appearance and geometric features reach higher accuracy performance, for example [14] and [15].

This paper proposes a fully automated FER system based on the combination of local Fourier coefficients (appearance features) and facial Fourier descriptors (geometric features) of independent-specific facial regions (eyes-eyebrows, nose and mouth). By performing independent subspaces on frequency domain for each facial region, we can approach common FER problems such as illumination changes, partial occlusion, image rotation, redundancy and dimensionality reduction. Finally, facial expressions are recognized using Support Vector Machines (SVMs) and evaluated with three widely used data sets: the Extended Cohn-Kanade (CK+) database [16], the Multimedia Understanding Group (MUG) database [17] and the Taiwanese Facial Expression Image Database (TFEID) [18].

This paper is strongly related to the work presented in [19]. However, it extends the previous work by introducing:

- A deeper literature review of similar works, which serves as comparison results.
- A fully automated fiducial point detection and region segmentation (instead

of manually annotated).

- A more detailed description of the method for making easier its reproduction.
- A complete evaluation using full-size data sets (not only culture-specific frames).
- A higher recognition rate performance obtained by getting the ideal number of fiducial points and size of sub-blocks.
- A study of the combinations of facial regions for facing the problem of partial occlusion.

In summary, the main contributions of this paper are:

- A fully automated FER system based on appearance and geometric features using local Fourier coefficients and SVMs.
- A study of local Fourier coefficients with different sizes of sub-blocks.
- A study of facial Fourier descriptors with a different number of fiducial points.
- A comparison results with the state-of-the-art methods facing the problem of partial occlusion.
- Extensive FER experiments on three different data sets demonstrating the efficiency of the proposed system above some previous works.

The rest of the paper is organized as follows: a review of related works is presented in Section 2. The general framework of the proposed FER system is explained in Section 3 followed by the description of data sets and the evaluation protocol in Section 4. Section 5 shows the experimental results and finally, the conclusion and future works are drawn in Section 6.

2. Related Works

Several studies have been proposed for combining the benefits of appearance and geometric features for FER. For instance, Li *et al.* [5] proposed a FER system which combines the appearance and shape information using bag of words and PHOG descriptors respectively. The authors applied SVMs for classifying both methods independently by using four component regions: forehead, eyes-eyebrows, nose and mouth. The combination of appearance and shape information was made at decision level, which implies that multi-class SVMs have to be applied twice before the final decision. Since it has to tune and train two different classifiers, this proposal presents several problems of computational complexity, therefore it is inefficient for real-time applications. However, the fusion at decision level allows the possibility to obtain independent results from only one kind of features. On the other hand, Yi *et al.* [20] proposed the feature combination at feature extraction level. They obtained a final feature vector from three facial features: feature point distance ratio (geometric), connection angle ratio (geometric) and skin deformation energy (appearance). It is worth noting that these features were obtained by taking the neutral frame as a baseline. Thus, both images (neutral and expressive) are required for the process. The skin deformation energy is calculated from a small region between eyebrows, which is defined by the fiducial points obtained with AAM. Lastly, in a more recent approach, Ghimire *et al.* [15] proposed a FER system based on LBP (appearance) and normalized central moments (geometric) of 29 specific local regions. Same as the previous proposal, for appearance features extraction, this method defines the local regions by employing fiducial points. Therefore, if there is a problem related to landmark detection, the appearance features could be strongly affected. In other words, the feasibility of appearance features depends on the good extraction of geometric features. In summary, for developing a robust FER system based on the fusion of appearance and geometric features, two issues have to be accounted: the computational complexity and the dependency of each feature extraction methods.

3. System Framework

The proposed FER system consists of four steps: face detection, facial region segmentation, feature extraction and classification. As shown in **Figure 1**, the first step is face detection which is performed by the well-known Viola-Jones algorithm. Subsequently, in order to reduce the dimensionality and to highlight some regions of interest, we segment the face in specific facial regions (facial region segmentation). Thus, the local regions of eyes-eyebrows, nose and mouth are segmented based on the relation of the eyes distance for appearance features. Along with fiducial points of the same regions for geometric features, which are estimated by the method proposed in [21]. Subsequently, the feature vector extraction is based on the combination of local Fourier coefficients (LFC) and facial Fourier descriptors (FFD), where each facial region represents an independent subspace based on PCA (principal component analysis). Also in this step, the concatenation of feature vectors of all facial regions is performed. Moreover, in order to overcome the identity bias, the final feature vector is represented as a linear difference of expressive and neutral information. Finally, the classification step is performed by SVM algorithm. SVMs are trained with feature vectors obtained from the combination of appearance and geometric features. It is worth noting that our proposal uses only one classifier even when three facial regions are involved in the feature extraction process, thus the computational complexity remains low. In addition, since the facial region segmentation of both kinds of features only depends on face detection, it can be considered that the feature extraction process is independent for both of them.

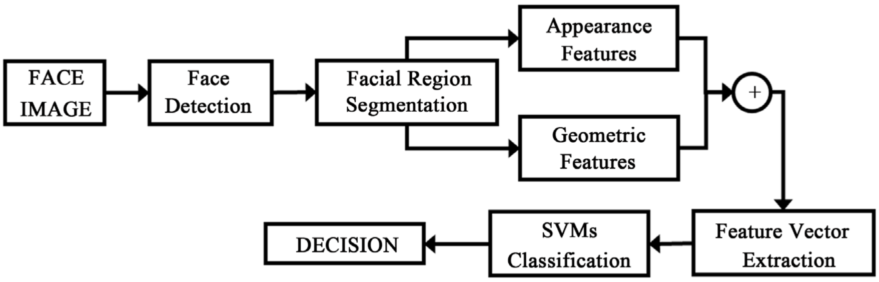

Figure 1. General framework of the proposed system.

3.1. Face Detection and Region Segmentation

As mentioned before, face detection is carried out by Viola-Jones algorithm. Thus, we obtain a detected face region (defined as ***DFR***) of size N^2, the eyes position of the face can be defined as E_L, E_R for left and right eye respectively, where $E_L, E_R \in \boldsymbol{DFR}$. Then, in order to segment the facial regions for appearance features, we used the distance between eyes, which is defined as $Di = |E_L - E_R|$ and experimentally we found the relation between Di and the three specific facial regions. For instance, consider O as the origin of the plane ***DFR***, where $O = (x_L - x_R, y_L - y_R)/2$. Thus, the upper left vertex of each facial region is defined as follows,

$$P_{Eye} = (x_O - Di, y_O + 2/5\,Di)$$

$$P_{Nos} = (x_O - 4/5\,Di, y_O - 2/5\,Di)$$

$$P_{Mou} = (x_O - 3/5\,Di, y_O - 4/5\,Di)$$

where $P_{Eye}, P_{Nos}, P_{Mou}$ represent the initial positions of eyes-eyebrows, nose and mouth regions respectively. Finally, the size of each facial region is defined as,

$$A_{Eye} = 2Di \cdot 4/5\,Di$$

$$A_{Nos} = 8/5\,Di \cdot 3/5\,Di$$

$$A_{Mou} = 6/5\,Di \cdot 4/5\,Di$$

being $A_{Eye}, A_{Nos}, A_{Mou}$ the area of the respective $\boldsymbol{FR}_{Eye}, \boldsymbol{FR}_{Nos}, \boldsymbol{FR}_{Mou}$ facial regions.

In order to obtain the fiducial points of each facial image we applied the work proposed in [21], where a deformable face tracking model was trained by employing a cascade of linear regression functions. The process consists of detecting the face in the first frame and then identifying the facial landmarks at each consecutive frame by using fitting results of the previous frame as initialization. This approach obtains 51 facial landmarks for describing the shapes of eyes-eyebrows (22), nose (11) and lips (18). This method has been tested for working under controlled environments as well as "in-the-wild" scenarios [22]. **Figure 2** illustrates an example of automatic segmentation for appearance and geometric features.

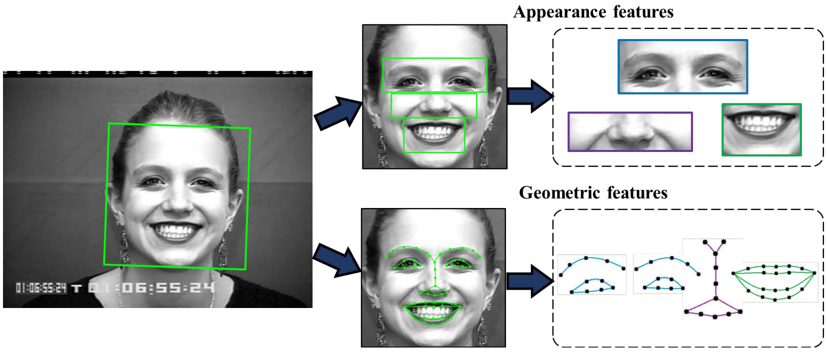

Figure 2. Example of face detection process and facial region segmentation for both kinds of features.

3.2. Feature Extraction

The basis of our proposal is the Fourier transform which has been applied a few times for facial recognition (FR) and FER. For instance, the method proposed in [23] fused three different Fourier feature domains for FR. On the other hand, the phase spectrum of FFT applied locally to non-overlapped sub-blocks has been proposed for FR [24] and FER [25] respectively. In addition, a method called LPQ (Local Phase Quantization) [26] defines local neighborhoods for obtaining local histograms of LPQ patterns (usually employing regions of 7 × 7 pixels) similarly to the LBP process, but instead of using pixel intensities, LPQ employs the phase of each neighborhood. However, those approaches employed only appearance features. Instead, this work proposes a combination of Local Fourier Coefficients (LFC) and Facial Fourier Descriptors (FFD). The feature extraction process consists of three steps: appearances feature extraction, geometric feature extraction and feature vector estimation. This process is applied independently for each facial region and each type of features. Hence, it could run in parallel if needed.

Appearance feature extraction is carried out by using LFC which builds on the 2-D DFT. This process consists of dividing the input image into several sub-blocks to locally extract Fourier coefficients. For instance, the 2-D DFT is defined as:

$$F(u,v)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\mathrm{e}^{-j2\pi(ux/M+vy/N)} \tag{1}$$

where $f(x,y)$ is a digital image of size $M \times N$ and it must be evaluated for values of the discrete variables u and v in the ranges of $u=0,1,2,\cdots,M-1$ and $v=0,1,2,\cdots,N-1$.

Consider $\boldsymbol{FR}_{roi}$ as the *roi*-th facial region image of size $M \times N$, and for convenience, ***FR*** represents any of the three facial regions which have to be divided into sub-blocks of size $L \times L$. Then, the local 2-D DFT of the current facial region is given by a modification of Equation (1):

$$f_{p,q}(u,v)=\sum_{x=0}^{L-1}\sum_{y=0}^{L-1} \boldsymbol{FR}_{p,q}(x,y)\mathrm{e}^{-j2\pi(ux/L+vy/L)} \tag{2}$$

where $0 \le u,v < L$, and $\boldsymbol{FR}_{p,q}(x,y)$ represents the (p,q)-th sub-block of the facial region $\boldsymbol{FR}$. Since the minimum sub-block size is $L=2$, the imaginary component of complex Fourier coefficients is equal to zero so that

$$f_{p,q}(u,v)=\boldsymbol{Re}(u,v)+j\times 0\times \boldsymbol{Im}(u,v) \tag{3}$$

where $\boldsymbol{Re}(u,v)$ and $\boldsymbol{Im}(u,v)$ are the real and imaginary components of $f_{p,q}(u,v)$ respectively. The ideal size of L has been analyzed in [24], however, this analysis is focused only on the phase component of the Fourier transform. Therefore, an analysis of the ideal sub-block size for real components of LFC is presented in Section 5.1.

Considering the Equation (3), the local Fourier coefficient matrix is given by:

$$\boldsymbol{lfc} = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,N/L} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,N/L} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M/L,1} & f_{M/L,2} & \cdots & f_{M/L,N/L} \end{bmatrix} \tag{4}$$

where ***lfc*** has the same dimensions as ***FR***. In summary, ***lfc*** matrix represents the real components of frequency features obtained locally by each sub-block of size $L \times L$.

Subsequently, a variation of PCA is applied in order to reduce the dimensionality and for correlating the local information with the set of training images. To this end, the ***lfc*** matrix is converted into a column vector, so that

$$\boldsymbol{V}_{lfc} = \mathrm{vec}\left(\boldsymbol{lfc}(m,n)\right) \tag{5}$$

where $\boldsymbol{V}_{lfc}$ is the column vector of ***lfc*** for $0 \le m,n < M,N$. Next, LFC vectors of the training set have to be concatenated to form the matrix $\boldsymbol{\Phi}$:

$$\boldsymbol{\Phi}_{lfc} = \left[\boldsymbol{V}_{lfc}^{0} - \boldsymbol{\mu}_{lfc}, \boldsymbol{V}_{lfc}^{1} - \boldsymbol{\mu}_{lfc}, \cdots, \boldsymbol{V}_{lfc}^{P-1} - \boldsymbol{\mu}_{lfc}\right] \tag{6}$$

where P is the total number of images used for training and $\boldsymbol{\mu}_{lfc}$ is the mean vector given by:

$$\boldsymbol{\mu}_{lfc} = \frac{1}{P}\sum_{n=0}^{P-1}\boldsymbol{V}_{lfc}(n) \tag{7}$$

Subsequently, the covariance matrix $\boldsymbol{\Omega}_{lfc}$ is estimated using the Equation (8), which is used to obtain P eigenvectors associated with non-zero eigenvalues, where $P < M \times N$.

$$\boldsymbol{\Omega}_{lfc} = \boldsymbol{\Phi}_{lfc}^{\mathrm{T}}\boldsymbol{\Phi}_{lfc} \tag{8}$$

Those eigenvectors are then stored in a descendent order according to the corresponding eigenvalues. The sorted eigenvectors of the covariance matrix determine the subspace $\boldsymbol{\Psi}_{lfc}$ associated to the current facial region, which is defined by

$$\boldsymbol{\Psi}_{lfc} = \left[\boldsymbol{V}_0^{\mathrm{T}}, \boldsymbol{V}_1^{\mathrm{T}}, \cdots, \boldsymbol{V}_{H-1}^{\mathrm{T}}\right] \tag{9}$$

where $\boldsymbol{V}_0$ is the eigenvector associated with the largest eigenvalue, $\boldsymbol{V}_1$ is the eigenvector associated with the second largest eigenvalue, and H is the number of eigenvectors used for further projections. It is worth noting that this process is applied so that 90% of the variance of training vectors is retained. Finally, the LFC feature vector $\boldsymbol{Y}_{lfc}$ is given by:

$$\boldsymbol{Y}_{lfc} = \boldsymbol{\Psi}_{lfc}^{\mathrm{T}}\left(\boldsymbol{V}_{lfc} - \boldsymbol{\mu}_{lfc}\right) \tag{10}$$

where $\boldsymbol{\Psi}_{lfc}$ is the facial region subspace and $\boldsymbol{\mu}_{lfc}$ is the mean vector of all training images.

On the other hand, geometric feature extraction process is based on FFD which uses Fourier Descriptors (FD). FFD represents a digital boundary of 1D Fourier coefficients estimated by a sequence of coordinate pairs transformed by applying the DFT. To this end, each facial region shape is considered as *K*-point coordinate pairs, K being the number of facial feature points of the shape. An analysis of the effect of different number of K is presented in Section 5.2.

For applying FFD, suppose that a specific shape of the $\boldsymbol{FR}$-th facial region is represented as a sequence of coordinates, so that

$$s_{FR}(k) = \left[x_{FR}(k), y_{FR}(k)\right] \tag{11}$$

where $k = 0,1,2,\cdots,K-1$. Afterwards, Equation (12) is applied to each coordinate pairs of the sequence, thus complex numbers are generated.

$$s(k) = \left[x(k) - x_c\right] + j\left[y(k) - y_c\right] \tag{12}$$

where (x_c, y_c) represents the centroid of the shape, which is the average of all coordinate pairs, so that

$$x_c = \frac{1}{K}\sum_{t=0}^{K-1} x(t), \quad y_c = \frac{1}{K}\sum_{t=0}^{K-1} y(t) \tag{13}$$

Subsequently, the FFD of $s(k)$ is given by.

$$ffd(u) = \sum_{k=0}^{K-1} s(k)\mathrm{e}^{-j2\pi uk/K} \tag{14}$$

for $u = 0,1,2,\cdots,K-1$, where $ffd(u)$ represents the Fourier Descriptors of the facial region shape, which have to be projected into the current facial region subspace similarly to the process of LFC. Therefore, the FFD feature vector $\boldsymbol{Y}_{ffd}$ is defined by using Equation (14) on the process described by Equations (5)-(9), thus:

$$\boldsymbol{Y}_{ffd} = \boldsymbol{\Psi}_{lfc}^{\mathrm{T}}\left(\boldsymbol{V}_{ffd} - \boldsymbol{\mu}_{ffd}\right) \tag{15}$$

The combination of both kinds of features comes at this point, where feature vectors of appearance and geometric features were individually calculated by Equations (10) and (15). The fusion begins with the concatenation of both feature vectors, so that

$$\boldsymbol{V}_{lfc+ffd} = \left[\boldsymbol{Y}_{lfc}^{\mathrm{T}}, \boldsymbol{Y}_{ffd}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{16}$$

Subsequently, the process of Equations (5)-(9) has to be applied once more. Thus, the final feature vector of one facial region is defined by:

$$\boldsymbol{Y}_{roi} = \boldsymbol{\Psi}_{lfc+ffd}^{\mathrm{T}}\left(\boldsymbol{V}_{lfc+ffd} - \boldsymbol{\mu}_{lfc+ffd}\right) \tag{17}$$

where $\boldsymbol{Y}_{roi}$ represents the feature vector of the $\boldsymbol{FR}_{roi}$ facial region . Therefore, a feature vector based on the three specific facial regions is defined as:

$$\boldsymbol{Y} = \bigcup_{l=1}^{C} \boldsymbol{Y}_{roi}(l), \tag{18}$$

where $\boldsymbol{Y}$ represents the concatenation of C individual facial regions. It is worth noting that C can be equal to 2 or 3 depending on how many facial regions are involved in the feature extraction process.

Finally, in order to overcome the identity bias, we follow the assumption that facial expressions can be represented as a linear combination of expressive and neutral face images of the same subject [27]. Therefore, as a final step, we propose to subtract final feature vectors from neutral and expressive images. Thus, the definitive feature vector is given by:

$$\boldsymbol{Z}(h)=\boldsymbol{Y}_{Exp}(h)-\boldsymbol{Y}_{Neu}(h) \tag{19}$$

for $h=0,1,2,\cdots,Q-1$, where Q is the total number of expressive images in the dataset, $\boldsymbol{Y}_{Exp}$ and $\boldsymbol{Y}_{Neu}$ represent the final feature vectors of expressive and neutral facial image, and $\boldsymbol{Z}$ the difference vector which is the definitive feature vector used in the classification stage.

3.3. Support Vector Machines Classification

Support Vector Machines (SVM) is an efficient classifier known for its generalization capability. Therefore, in this paper, a multi-class SVMs employing radial basis function (RBF) kernels were used in order to classify the six basic facial expressions. The library LIBSVM [28] is used to achieve this task. SVM has to be applied in two different modalities: training and testing. Thus, a set of feature vectors should be introduced to the classifier as training images. Accordingly, by maximizing the hyperplane margin, the SVMs obtain six templates, which are linked to the facial expressions of anger, disgust, fear, happiness, sadness, and surprise. Afterward on the testing mode, in broad outlines, the SVMs compare the test feature vector with all templates to decide from which class it belongs to. It is important to mention that this decision depends on the facial region combination gotten by the previous stage.

4. Data Sets and Evaluation Protocol

A subset of the Extended Cohn-Kanade (CK+) database [16] was selected for the analysis of sub-block size and the number of facial landmarks (for LFC and FFD respectively). It includes expressive and neutral faces of 90 different subjects. In order to avoid misinterpretations of the results due to the data set, this subset was selected by choosing the same number of expressive images per basic facial expression (40 images). Hence, 240 peak expressive frames and 90 neutral frames (from each subject) were selected from the available sequences of CK+.

The fully automated system was evaluated using the complete version of the CK+ database, the Multimedia Understanding Group (MUG) database [17] and the Taiwanese Facial Expression Image Database (TFEID) [18]. **Table 1** shows the number of subjects and frames per expression of each data set, where 362 expressive frames comprise the CK+, 304 the MUG and 229 the TFEID. It is important to mention that for CK+ data set, the number of images from the expressions of fear and sadness was increased by selecting two expressive frames

Table 1. Number of images and subject of each data set.

Data Set	Ang.	Dis.	Fea.	Hap.	Sad.	Sur.	Subjects
CK+	45	59	**50**	69	**56**	83	116
MUG	52	51	48	52	49	52	52
TFEID	34	40	40	40	39	36	40

from each sequence (not only peak frames). Thus, the original number of sequences of these expressions is 25 and 28 respectively.

The system was evaluated following a widely used protocol in FER, this is leave-one-subject-out (LOSO) cross-validation. This method consists of dividing the database according to the number of subjects, such as each sub-group consists of only images from the same subject. Then, one of these sub-groups has to be picked out for testing and the remaining are used for training. This procedure has to be repeated the same number of times as the number of subjects in the database. Finally, the recognition accuracy is averaged over all trials. In addition to the average recognition rate of LOSO, confusion matrices are also presented for evaluation results. The diagonal entries of the confusion matrices represent the accuracy of the facial expressions correctly classified, whereas the off-diagonal rates the misclassification problems.

5. Experimental Results

The experimental results are divided into four main tests: analysis of sub-block sizes for LFC, where different sizes of sub-blocks are tested using the subset of CK+; analysis of the number of landmarks for FFD, where the geometric features are defined with different number of landmarks using the same subset as the previous test; results of LFC + FFD with all data sets, this test presents the results of the main proposal of this paper using CK+, MUG and TFEID; and the comparison with previous methods presents the performance of different approaches which used the same data sets, a comparison with approaches that overcome partial occlusions are also presented in this test.

5.1. Analysis of Sub-Block Size for LFC

In this section, several variations of sub-block sizes are proposed in order to find the ideal sub-block size for LFC. Based on the analysis presented in [24], the ideal sub-block size of the Eigenphases algorithm is the minimum possible window (*i.e.* 2 × 2 pixels). However, Eigenphases employs the phase spectrum instead of Fourier coefficients as we proposed for LFC. Therefore, by adopting the analysis bias of [24] we test the LFC method with four square sizes ($L = 2$, $L = 4$, $L = 6$ and $L = 12$). In addition, three non-square windows are proposed: $L = \text{M·N}/2$, $L = \text{M}/3$ and $L = \text{N}/3$, which represent the segmentation of the facial region into four and three equal size parts (horizontal and vertical possibilities). Finally, the whole input facial region ($L = \text{M·N}$) without local segmentation is also tested. Figure 3 illustrates an example of sub-block region segmentation of the described non-square windows applied to the mouth facial region. In sum-

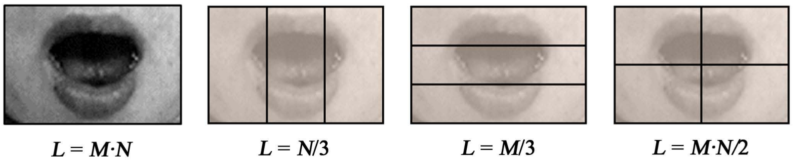

Figure 3. Examples of sub-block segmentation of non-square regions.

mary, the following analysis presents the performance of LFC when eight different sizes of L in Equation (2) are used for feature vector calculation.

The results of the eight different sub-block sizes are shown in Figure 4. It is worth noting that these were obtained using a subset of CK+ which has the same number of images per basic expression, and the evaluation performance was as described in Section 4. From this graph we can easily see that the best recognition performance is obtained using the combination of all facial regions in the feature extraction process. In addition, as presented in [24] the average recognition rate increases when the size of the sub-block decreases. Thus, the best results are obtained by using $L = 2$ which represents the minimum square window of just 2 × 2 pixels. Finally, we can highlight that the best performance of LFC is reached when the sub-block size is equal to 2 × 2 pixels.

5.2. Analysis of the Number of Fiducial Points for FFD

Choosing the number of landmarks that defines the facial shape is an important issue for every FER system based on geometric features. Therefore, a test for FFD using eight different number of fiducial points ($K = 31, 41, 51, 64, 81, 93, 115$ and 123) is presented in this section. The test consists of analyzing FER performance based on different shape representations by changing the number of landmarks used in Equation (11). It is important to mention that for this particular test, the landmark estimation was manually annotated for all images of the CK+ subset. The main differences between the eight shape representations reside on the location and the number of facial landmarks of each facial region. For example, for $K = 31$ the number of landmarks representing the nose region is 7 whereas for $K = 123$ the same region is represented by 29 landmarks. Figure 5 shows three examples of these different shape representations, *i.e.* $K = 31$, $K = 51$, and $K = 123$.

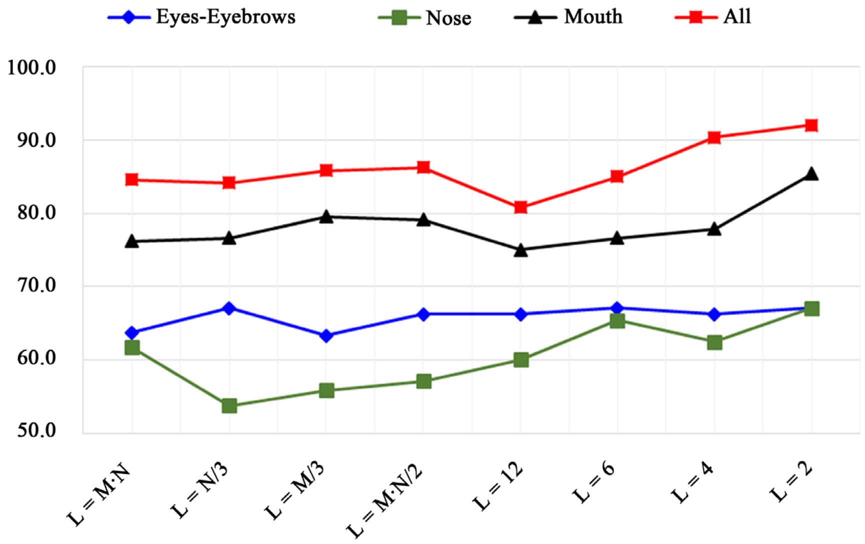

Figure 4. Results of LFC with different sub-block sizes using eyes-eyebrows, nose, mouth and all facial regions for feature extraction process.

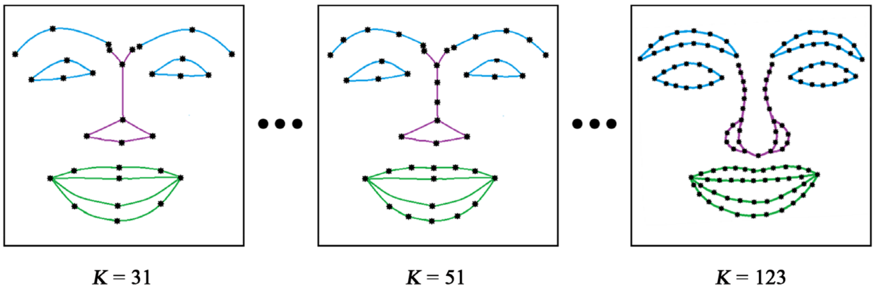

Figure 5. Examples of facial shapes represented by different number of fiducial points (K).

Results of the eight K values for FFD are shown in Figure 6. This figure presents individual performance of eyes-eyebrows, nose lips and the combination of all of them. As expected, we can see that the results improve when the number of landmarks increases, thus $K = 123$ presents the best performance for FFD. However, the improvement is not significant for some tests. For example, when all regions are used for feature extraction (All) the average recognition rates of $K = 51$ and $K = 123$ are 93.8% and 95.9% respectively, just 2% of improvement. Moreover, even when the nose region presents a remarkable improvement of accuracy, this is not reflected when all the regions are used for the feature extraction. Therefore, we decided to use the number of landmarks provided by [21] which conveniently is $K = 51$.

Finally, the last test performed with the subset of CK+ is a comparison of LFC, FFD and LFC + FFD methods using the ideal size of sub-block and the chosen number of facial landmarks ($L = 2$ and $K = 51$ respectively). It is important to mention that final feature vectors of LFC, FFD and LFC + FFD were obtained using the combination of all facial regions, as defined in Equations (10), (15) and (17) respectively. Table 2 presents the results of individual and all possible facial regions, which shows that LFC + FFD obtains higher accuracy than individual LFC and FFR. In turn, the results of geometric features (FFD) slightly overcome those of appearance features (LFC).

5.3. Results of LFC + FFD for All Data Sets

This section presents the results of our main proposal, the fully automated FER system based on LFC + FFD. Feature vectors were obtained with Equation (19) and classified by SVMs as described in Section 3.3. Results obtained with full data sets of CK+, MUG, and TFEID are presented in Table 3. From this table, we can see that the best performance among all data sets is reached by using all regions for feature extraction (All regions). Moreover, the best results using two and one facial regions are based on Eyes-Eyebrows-Mouth and Mouth respectively, for all data sets. Indeed, the performance of using only two facial regions is highly competitive, only approximately 1% of accuracy is decreased compared with "All regions". On the other hand, the results with CK+ present a wider gap of the average recognition rate (10%) between Mouth and Eyes-Eyebrows regions. Furthermore, the TFEID test presents a significant decrease of perfor-

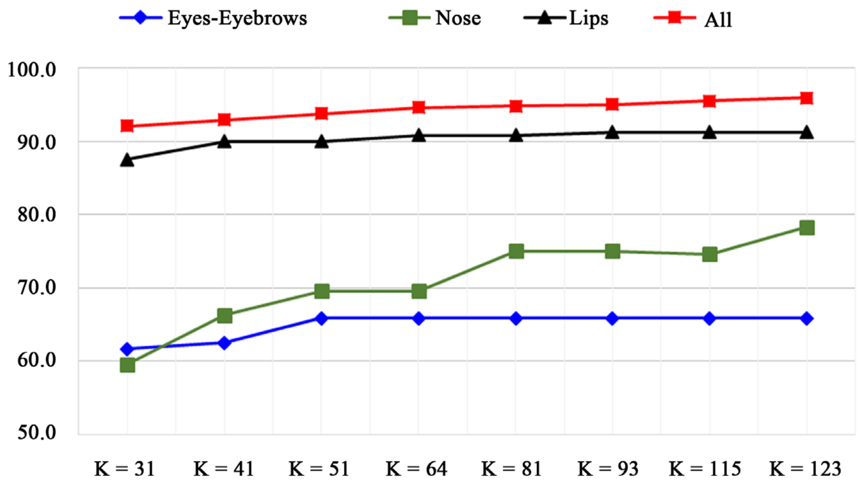

Figure 6. Results of FFD with different number of fiducial points using eyes-eyebrows, nose, mouth and all facial regions for feature extraction process.

Table 2. Classification accuracy (%) of proposed LFC, FFD and LFC + FFD methods with ideal sizes of $L = 2$ and $K = 51$. Performance based on individual facial regions and its combinations.

Method:	LFC	FFD	LFC + FFD
Eyes-Eyebrows	67.1	65.8	71.7
Nose	67.1	69.6	75.8
Mouth	85.4	90.0	90.8
Eyes-Eyebrows-Nose	76.7	66.7	78.8
Eyes-Eyebrows-Mouth	90.0	92.1	94.2
Nose-Mouth	89.2	85.0	93.3
All Regions	92.1	92.5	95.8

Table 3. Classification accuracy (%) of the fully automated FER system (LFC + FFD) evaluated with three standard data sets. Performance based on individual facial regions and its combinations.

Data Set:	CK+	MUG	TFEID
Eyes-Eyebrows	78.7	81.1	77.8
Nose	86.2	80.2	74.7
Mouth	87.7	85.7	80.4
Eyes-Eyebrows-Nose	89.8	88.5	86.7
Eyes-Eyebrows-Mouth	96.4	94.0	93.0
Nose-Mouth	94.0	89.9	88.0
All Regions	97.9	95.9	94.9

mance when less than two regions are used for feature extraction. In other words, it is more difficult to recognize the six basic expressions using only one facial region with TFEID data set. In summary, the best performance reached by our proposal is based on all regions for feature extraction and the mouth seems to be the facial region which can better represent the six basic expressions.

Tables 4-6 present the confusion matrices of the proposed system evaluated with CK+, MUG and TFEID respectively. From these tables, we can see that among all data sets, sadness is the expression with higher recurrent misclassification problems, and in most of the cases this expression is misrecognized with anger (3%, 10% and 14% for CK+, MUG and TFEID respectively). On the other hand, surprise is the only expressions which obtain a perfect recognition accuracy among all data sets. Some specific situations are presented for each data set, e.g. for CK+ anger is misrecognized with disgust; for MUG disgust is misrecognized with happiness; and for TFEID fear is misrecognized with disgust. In general, it can be summarized that the expressions of surprise and happiness are the easiest to recognize among all basic expressions, whereas sadness and anger are the most difficult.

Table 4. Confusion matrix using LFC + FFD on CK+ data set.

	Ang.	Dis.	Fea.	Hap.	Sad.	Sur.
Anger	**95.7**	2.2	0	0	2.2	0
Disgust	1.7	**98.3**	0	0	0	0
Fear	0	0	**95.7**	4.3	0	0
Happiness	0	0	1.5	**98.5**	0	0
Sadness	3.0	0	0	0	**97.0**	0
Surprise	0	0	0	0	0	**100**

Table 5. Confusion matrix using LFC + FFD on MUG data set.

	Ang.	Dis.	Fea.	Hap.	Sad.	Sur.
Anger	**94.6**	0	2.7	0	2.7	0
Disgust	0	**97.0**	0	3.0	0	0
Fear	0	0	**100**	0	0	0
Happiness	0	0	0	**100**	0	0
Sadness	10.3	3.4	6.9	0	**79.3**	0
Surprise	0	0	0	0	0	**100**

Table 6. Confusion matrix using LFC + FFD on TFEID data set.

	Ang.	Dis.	Fea.	Hap.	Sad.	Sur.
Anger	**91.7**	0	4.2	0	4.2	0
Disgust	0	**100**	0	0	0	0
Fear	0	4.2	**95.8**	0	0	0
Happiness	0	0	2.7	**97.3**	0	0
Sadness	14.3	4.8	0	0	**81.0**	0
Surprise	0	0	0	0	0	**100**

5.4. Comparison with Previous Methods

A comparison with other approaches evaluated with same data sets is shown in this section. CK+ is one of the most used data sets for FER, therefore Table 7 presents just some of many approaches which have employed it. From this table, we can see that our proposal overcomes all previous approaches. However, works [14], [15] and [27] also present an average recognition rate higher than 97%. It is worth noting that two of these approaches used a combination of appearances and geometric features. In general, it can be noticed that the approaches based on both kinds of features reach higher performance. In addition, our proposal also overcomes results obtained by approaches based on Deep Neural Networks [9] [14], which provide semantic features of expressive faces.

Table 8 and Table 9 present the comparison of performance of different approaches with MUG and TFEID respectively. In both cases, our proposal obtains

Table 7. Comparison with different approaches with CK+.

Ref. & Year	Method	Classifier	Data	Features	Protocol	Accuracy (%)
[20] ‘14	FPDRC + CARC + SDEP	NN	Image	Both	-	88.70
[29] ‘16	Weighted Feats.	SVM	Image	Geo.	2-fold	93.00
[8] ‘09	Boosted LBP	SVM	Image	App.	10-fold	95.10
[12] ‘11	PCA	LDCRF	Sequence	Geo.	4-fold	95.79
[10] ‘15	DVNP	RF	Sequence	Geo.	10-fold	96.38
[9] ‘17	CNN	LR	Image	App.	8-fold	96.76
[27] ‘14	PCA Dictionary	SRC	Image	App.	LOSO	97.19
[15] ‘16	LBP + NCM	SVM	Image	Both	5-fold	97.25
[14] ‘15	CNN + DNN	Joint F-N	Sequence	Both	10-fold	97.25
Proposed	LFC + FFD	SVM	Image	Both	LOSO	**97.90**

a. “App.” and “Geo.” refer to appearance and geometric features respectively; b. “Both” refers to a combination of appearance and geometric features.

Table 8. Comparison with different approaches with MUG.

Ref. & Year	Method	Classifier	Data	Features	Protocol	Accuracy (%)
[30] ‘15	Gabor + PCA	NN	Image	App.	2-fold	89.29
[13] ‘16	Landmark Dist.	SVM	Image	Geo.	2-fold	90.50
[7] ‘13	LFDA	kNN	Image	App.	LOSO	95.24
[11] ‘17	Triangle Land.	SVM	Sequence	Geo.	10-fold	95.50
Proposed	LFC + FFD	SVM	Image	Both	LOSO	**95.85**

Table 9. Comparison with different approaches with TFEID.

Ref. & Year	Method	Classifier	Data	Features	Protocol	Accuracy (%)
[31] ‘17	Haar Wavelet	LR	Image	App.	10-fold	89.58
[32] ‘14	LBP + MPC	SVM	Image	App.	10-fold	92.54
[33] ‘17	Pyramid Feat.	SVM	Image	App.	LOSO	93.38
[34] ‘15	DSNGE	kNN	Image	App.	LOSO	93.89
Proposed	LFC + FFD	SVM	Image	Both	LOSO	**94.94**

the highest recognition accuracy. This occurs, even when some approaches don't use the complete data set of MUG, like [7], and the process is based on sequence of frames, as in [11]. It is worth noting that the TFEID data set presents a bigger challenge for FER because instead of CK+ and MUG, the facial expressions are shown only by Taiwanese people. Therefore, the face structure and some facial expressions may be affected by cultural differences.

The last comparison with different approaches is focused on the capability to handle the partial occlusion problem. Methods [6], [25], [28] and [35] proposed different approaches for solving this problem. Our potential solution consists of excluding the occluded facial region in the feature extraction process. For example, for eyes-eyebrows occlusion, our system only uses the regions of mouth and nose for feature vector estimation. **Table 10** compares the results of methods under the occlusion of a specific facial region. In this situation, our proposal presents competitive results with other approaches. However, those are based on CK data set which is a previous version of CK+ known to be limited in size and lacked of spontaneous and non-exaggerated expression. On the other hand, **Table 11** presents an opposite situation, *i.e.* when only one part of the face is available because of occlusion problems. This extreme case is approached for only a few methods, such as [25] and [28]. From this table we can see that our proposal presents higher recognition rates for each extreme situation. In addition, it can be noticed that the recognition performance is higher when the mouth is available. Therefore, the most difficult scenario related to partial occlusion is when the mouth region is occluded. In this situation, our system can reach 89.8% of accuracy if eyes-eyebrows and nose regions are available.

Table 10. Comparison with different approches under partial occlusion of specific facial regions.

				Occluded Part (%)		
Ref. & Year	Data Set	Method	Classifier	Eyes	Mouth	NO
[25] '14	CK	Eigenphases	SVM	87.7	75.3	92.0
[35] '12	CK	Random Gabor Filters	SVM	90.5	82.9	91.5
[6] '14	CK	Radial Gabor Filters	LDA + kNN	**95.1**	**90.8**	95.3
Proposed	CK+	LFC + FFD	SVM	94.0	89.8	**97.9**

a. "NO" refers to No Occlusion.

Table 11. Comparison with different approaches which present results with only one facial region.

				One Region Test (%)		
Ref. & Year	Data Set	Method	Features	Eyes	Nose	Mouth
[29] '16	CK+	Weighted Feats.	Geo.	41.9	25.5	60.4
[25] '14	CK	Eigenphases	App.	53.3	61.0	79.3
Proposed	CK+	LFC + FFD	Both	**78.7**	**86.2**	**87.7**

6. Conclusions and Future Work

In this paper, we proposed a fully automated FER system based on the combination of two novel feature extraction methods: LFC and FFD, which are focused on appearance and geometric features obtained from individual facial regions of eyes-eyebrows, nose and mouth. Therefore, our proposal is robust to common FER problems such as illumination changes, image rotation and dimensionality reduction. In addition, more different than the reviewed state-of-the-art approaches, our proposal could work well even when fiducial points are not accurately detected. This is possible because the appearance feature extraction does not depend on the extraction of geometric features. Thus, this proposal just depends on face and eyes detection, carried out by the robust algorithm of Viola-Jones, which achieved 100% of recognition with all data sets tested. Evaluation results also show that the proposed system can handle problems of partial occlusion without heavily decreasing its accuracy performance.

In general, results obtained with the proposed algorithm overcome most of the previous works. In addition, compared with recently famous methods such as CNN and DNN, our system shows better performance with CK+, MUG and TFEID data sets, reaching 98%, 96% and 95% respectively. On the other hand, we admit that the present work could present some limitations based on possible problems of head pose variations and non-frontal images. Therefore, in order to efficiently recognize spontaneous facial expressions, we will focus on solving these problems as a future work.

Finally, the proposed method should be also valid in other applications such as face recognition, facial action unit recognition and facial image understanding. Therefore, in the future, we would like to apply our method in some of these possible applications.

Acknowledgements

We would like to thank the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan for the Japanese government (Monbukagakusho) scholarship which supports the Ph.D. studies of the first author.

References

[1] Ekman, P. (1972) Universal and Cultural Differences in Facial Expression of Emotion. *Proceeding of Symposium on Motivation*, Nebraska University Press, Lincoln, Nebraska, **19**, 9-15.

[2] Tian, Y., Kanade, T. and Cohn, J.F. (2011) Facial Expression Recognition. In: Li, S.Z. and Jain, A.K., Eds., *Handbook of Face Recognition*, Springer, London, 487-519. https://doi.org/10.1007/978-0-85729-932-1_19

[3] Sariyanidi, E., Gunes, H. and Cavallaro, A. (2015) Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 1113-1133. https://doi.org/10.1109/TPAMI.2014.2366127

[4] Deshmukh, S., Patwardhan, M. and Mahajan, A. (2016) Survey on Real-Time Facial Expression Recognition Techniques. *IET Biometrics*, **6**, 155-163.

[5] Li, Z., Imai, J.I. and Kaneko, M. (2010) Facial Expression Recognition Using Facial-Component-Based Bag of Words and PHOG Descriptors. *Journal of ITE*, **64**, 230-236. https://doi.org/10.3169/itej.64.230

[6] Gu, W., Xiang, C., Venkatesh, Y.V., Huang, D. and Lin, H. (2012) Facial Expression Recognition Using Radial Encoding of Local Gabor Features and Classifier Synthesis. *Pattern Recognition*, **45**, 80-91. https://doi.org/10.1016/j.patcog.2011.05.006

[7] Rahulamathavan, Y., Phan, R.C.W., Chambers, J.A. and Parish, D.J. (2013) Facial Expression Recognition in the Encrypted Domain Based on Local Fisher Discriminant Analysis. *IEEE Transactions on Affective Computing*, **4**, 83-92. https://doi.org/10.1109/T-AFFC.2012.33

[8] Shan, C., Gong, S. and McOwan, P.W. (2009) Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study. *Image and Vision Computing*, **27**, 803-816. https://doi.org/10.1016/j.imavis.2008.08.005

[9] Lopes, A.T., de Aguiar, E., De Souza, A.F. and Oliveira-Santos, T. (2017) Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order. *Pattern Recognition*, **61**, 610-628. https://doi.org/10.1016/j.patcog.2016.07.026

[10] Pu, X., Fan, K., Chen, X., Ji, L. and Zhou, Z. (2015) Facial Expression Recognition from Image Sequences Using Twofold Random Forest Classifier. *Neurocomputing*, **168**, 1173-1180. https://doi.org/10.1016/j.neucom.2015.05.005

[11] Ghimire, D., Lee, J., Li, Z.N. and Jeong, S. (2017) Recognition of Facial Expressions Based on Salient Geometric Features and Support Vector Machines. *Multimedia Tools and Applications*, **76**, 7921-7946. https://doi.org/10.1007/s11042-016-3428-9

[12] Jain, S., Hu, C. and Aggarwal, J.K. (2011) Facial Expression Recognition with Temporal Modeling of Shapes. *Proceedings of* 2011 *IEEE International Conference on Computer Vision Workshops* (*ICCV Workshops*), 6-13 November 2011, Barcelona, 1642-1649. https://doi.org/10.1109/ICCVW.2011.6130446

[13] Maximiano da Silva, F.A. and Pedrini, H. (2016) Geometrical Features and Active Appearance Model Applied to Facial Expression Recognition. *International Journal of Image and Graphics*, **16**, 17. https://doi.org/10.1142/S0219467816500194

[14] Jung, H., Lee, S., Yim, J., Park, S. and Kim, J. (2015) Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 2983-2991. https://doi.org/10.1109/ICCV.2015.341

[15] Ghimire, D., Jeong, S., Lee, J. and Park, S.H. (2016) Facial Expression Recognition Based on Local Region Specific Features and Support Vector Machines. *Multimedia Tools and Applications*, **76**, 7803-7821. https://doi.org/10.1007/s11042-016-3418-y

[16] Kanade, T., Cohn, J.F. and Tian, Y. (2000) Comprehensive Database for Facial Expression Analysis. *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition* (*FG* 2000), Washington DC, 28-30 March 2000, 46-53. https://doi.org/10.1109/AFGR.2000.840611

[17] Aifanti, N., Papachristou, C. and Delopoulos, A. (2010) The MUG Facial Expression Database. *Proceedings of the* 11*th International Workshop on Image Analysis for Multimedia Interactive Services* (*WIAMIS*), Desenzano, 12-14 April 2010, 1-4.

[18] Chen, L.F. and Yen, Y.S. (2007) Taiwanese Facial Expression Image Database. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei.

[19] Benitez-Garcia, G., Nakamura, T. and Kaneko, M. (2017) Methodical Analysis of Western-Caucasian and East-Asian Basic Facial Expressions of Emotions Based on Specific Facial Regions. *Journal of Signal and Information Processing*, **8**, 78-98.

https://doi.org/10.4236/jsip.2017.82006

[20] Yi, J., Mao, X., Chen, L., Xue, Y. and Compare, A. (2014) Facial Expression Recognition Considering Individual Differences in Facial Structure and Texture. *IET Computer Vision*, **8**, 429-440. https://doi.org/10.1049/iet-cvi.2013.0171

[21] Asthana, A., Zafeiriou, S., Cheng, S. and Pantic, M. (2014) Incremental Face Alignment in The Wild. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 1859-1866.

[22] Chrysos, G.G., Antonakos, E., Snape, P., Asthana, A. and Zafeiriou, S. (2016) A Comprehensive Performance Evaluation of Deformable Face Tracking "In-the-Wild". *International Journal of Computer Vision*, 1-35. https://doi.org/10.1007/s11263-017-0999-5

[23] Hwang, W., Wang, H., Kim, H., Kee, S.C. and Kim, J. (2011) Face Recognition System Using Multiple Face Model of HYBRID FOURIER FEATURE under Uncontrolled Illumination Variation. *IEEE Transactions on Image Processing*, **20**, 1152-1165. https://doi.org/10.1109/TIP.2010.2083674

[24] Benitez-Garcia, G., Olivares-Mercado, J., Sanchez-Perez, G., Nakano-Miyatake, M. and Perez-Meana, H. (2013) A Sub-Block-Based Eigenphases Algorithm with Optimum Sub-Block Size. *Knowledge-Based Systems*, **37**, 415-426. https://doi.org/10.1016/j.knosys.2012.08.023

[25] Benitez-Garcia, G., Sanchez-Perez, G., Perez-Meana, H., Takahashi, K. and Kaneko, M. (2014) Facial Expression Recognition Based on Facial Region Segmentation and Modal Value Approach. *IEICE Transactions on Information and Systems*, **97**, 928-935. https://doi.org/10.1587/transinf.E97.D.928

[26] Rahtu, E., Heikkilä, J., Ojansivu, V. and Ahonen, T. (2012) Local Phase Quantization for Blur-Insensitive Image Analysis. *Image and Vision Computing*, **30**, 501-512. https://doi.org/10.1016/j.imavis.2012.04.001

[27] Mohammadi, M.R., Fatemizadeh, E. and Mahoor, M.H. (2014) PCA-Based Dictionary Building for Accurate Facial Expression Recognition via Sparse Representation. *Journal of Visual Communication and Image Representation*, **25**, 1082-1092. https://doi.org/10.1016/j.jvcir.2014.03.006

[28] Chang, C.C. and Lin, C.J. (2011) LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology* (*TIST*), **2**, 1-27. https://doi.org/10.1145/1961189.1961199

[29] Wei, W. and Jia, Q. (2016) Weighted Feature Gaussian Kernel SVM for Emotion Recognition. *Computational Intelligence and Neuroscience*, **2016**, 11-17. https://doi.org/10.1155/2016/7696035

[30] da Silva, F.A.M. and Pedrini, H. (2015) Effects of Cultural Characteristics on Building an Emotion Classifier through Facial Expression Analysis. *Journal of Electronic Imaging*, **24**, Article ID: 0230151. https://doi.org/10.1117/1.JEI.24.2.023015

[31] Goyani, M. and Patel, N. (2017) Multi-Level Haar Wavelet Based Facial Expression Recognition Using Logistic Regression. *Indian Journal of Science and Technology*, **10**, 1-9. https://doi.org/10.17485/ijst/2017/v10i9/108944

[32] Farajzadeh, N., Pan, G. and Wu, Z. (2014) Facial Expression Recognition Based on Meta Probability Codes. *Pattern Analysis* & *Applications*, **17**, 763-781. https://doi.org/10.1007/s10044-012-0315-5

[33] Ashir, A.M. and Eleyan, A. (2017) Facial Expression Recognition Based on Image Pyramid and Single-Branch Decision Tree. *Signal, Image and Video Processing*, 1-8. https://doi.org/10.1007/s11760-016-1052-9

[34] Kung, H.W., Tu, Y.H. and Hsu, C.T. (2015) Dual Subspace Nonnegative Graph

Embedding for Identity-Independent Expression Recognition. *IEEE Transactions on Information Forensics and Security*, **10**, 626-639. https://doi.org/10.1109/TIFS.2015.2390138

[35] Zhang, L., Tjondronegoro, D. and Chandran, V. (2014) Random Gabor Based Templates for Facial Expression Recognition in Images with Facial Occlusion. *Neurocomputing*, **145**, 451-464. https://doi.org/10.1016/j.neucom.2014.05.008