

Methodical Analysis of Western-Caucasian and East-Asian Basic Facial Expressions of Emotions Based on Specific Facial Regions

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.ucc.ac.jp

How to cite this paper: 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>

Received: March 30, 2017

Accepted: May 15, 2017

Published: May 18, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Facial expressions are the straight link for showing human emotions. Psychologists have established the universality of six prototypic basic facial expressions of emotions which they believe are consistent among cultures and races. However, some recent cross-cultural studies have questioned and to some degree refuted this cultural universality. Therefore, in order to contribute to the theory of cultural specificity of basic expressions, from a composite viewpoint of psychology and HCI (Human Computer Interaction), this paper presents a methodical analysis of Western-Caucasian and East-Asian prototypic expressions focused on four facial regions: forehead, eyes-eyebrows, mouth and nose. Our analysis is based on facial expression recognition and visual analysis of facial expression images of two datasets composed by four standard databases CK+, JAFFE, TFEID and JACFEE. A hybrid feature extraction method based on Fourier coefficients is proposed for the recognition analysis. In addition, we present a cross-cultural human study applied to 40 subjects as a baseline, as well as one comparison of facial expression recognition performance between the previous cross-cultural tests from the literature. With this work, it is possible to clarify the prior considerations for working with multicultural facial expression recognition and contribute to identifying the specific facial expression differences between Western-Caucasian and East-Asian basic expressions of emotions.

Keywords

Facial Expression Recognition, Cultural Specificity of Facial Expressions, Universality of Emotions, Cross-Cultural Analysis, Discrete Fourier Transform

1. Introduction

Facial expressions are a set of facial muscle movements which can directly ex-

press human emotions. Since the early work of Charles Darwin [1], facial expressions have been considered as a universal language, which can be recognized across different races and cultures around the world. Paul Ekman *et al.* [2] established the universality of basic facial expressions of emotions which are consistent among cultures. Thus, the prototypic expressions of anger, disgust, fear, happiness, sadness and surprise have known to be universally recognized and expressed. In order to methodically measure them, the Emotional Facial Action Coding System (EMFACS) was proposed [3]. EMFACS focuses on muscle movements called Action Units (AUs), which are related to the six basic emotions, and it provides specific coding for each of them. On the other hand, a certain part of the research community has promoted an opposite theory of the universality hypothesis. They proposed that the facial expressions are based on cultural learning and different races may have different ways to express emotions [4]. Indeed, recent cross-cultural studies have to some degree refuted this assumed universality by finding differences on facial expressions of Western-Caucasian and East-Asian people [5] [6].

In spite of the psychological findings on the cultural specificity of facial expressions, from the viewpoint of Human-Computer Interaction (HCI) the cultural universality of basic expressions is taken for granted [7]. Thus, automatic facial expression recognition systems (FER) do not consider the cultural differences of facial expressions [8]. Therefore, in order to develop more realistic HCI systems, the differences of facial expressions among cultures have to be taken into account by FER systems. Furthermore, a deep analysis of recognition and representation of those differences is needed. Some of the practical applications of the results of this analysis include the development of highly accurate human-robot interaction systems for social robots and that of culturally specific avatars.

The general approach of any FER system is based on three steps: face detection, feature extraction and classification [9]. In order to make an accurate analysis of facial expressions, the feature extraction process is crucial. Some of the techniques which are successfully applied to feature extraction are Bag of Words [10], Gabor [11], LBP [12], HOG [13] and recently Convolutional Neural Networks (CNN) [14]. On this category, Fourier transform has accurately faced the problems of illumination changes, image rotation and dimensionality reduction of face recognition and facial expression recognition systems [15] [16] [17].

In order to contribute to the theory of cultural specificity of basic expressions of emotions, from a composite viewpoint of psychology and HCI, this paper presents a methodical analysis of Western-Caucasian and East-Asian prototypic facial expressions based on four specific facial regions: forehead, eyes-eyebrows, mouth and nose. Our analysis is composed by facial expression recognition and visual analysis of facial expression images from four standard databases CK+, JAFFE, TFEID and JACFE, divided in two datasets of different cultural and ethnic regions: Western-Caucasian (WSN) and East-Asian (ASN). All possible testing-training combinations of these datasets plus a composition of both (mul-

multicultural dataset) were analyzed. For FER analysis, we proposed a hybrid feature extraction method based on Fourier coefficients calculated from geometric and appearance features which are classified by Support Vector Machines (SVMs). In order to identify facial differences of the six basic expressions, we present two independent visual representations: reconstructed images from Eigenfaces using appearance features [18]; and caricature faces using geometric features [19]. In addition to this proposed analysis, we present a human study applied to 40 subjects from WSN and ASN ethnic races as a baseline.

In summary, the main contributions of this paper include supplementary findings to the theory of cultural specificity for FER based on four individual facial regions and its combinations, as well as a visual analysis of the differences between Western-Caucasian and East-Asian basic facial expressions. Thus, the prior considerations for working with multicultural FER are also presented. Furthermore, the proposed analysis is based on a novel feature extraction method which combines the most relevant geometric and appearance features of the facial expression. This method employs the coefficients of the 2-D Discrete Fourier Transform and Fourier Descriptors for each type of feature respectively.

The rest of the paper is organized as follows: a review of related works is presented in Section 2. The proposed FER system is explained in Section 3 followed by the visual analysis proposal described in Section 4. Section 5 shows the experimental results and finally, the conclusion and future works are drawn in Section 6.

2. Related Works

Related works are clearly divided by psychological and HCI viewpoints. Psychological studies try to prove the refutation of the universality hypothesis of facial expressions. Meanwhile, those of HCI attempt to explain the low accuracy performance obtained by cross-cultural tests, taking the universality hypothesis as granted. From the psychological viewpoint, the works of Dailey *et al.* [5] and Jack *et al.* [6] present similar analysis based on human studies using cross-cultural datasets as stimuli. They analyzed the effect of culture-specific facial expression recognition capability of U.S. (Westerns) and Japanese participants (East-Asians). The authors found in-group advantages for recognizing facial expressions and they concluded that the 6 basic emotions can clearly represent the Western facial expressions, but they are inadequate to accurately represent the East-Asians emotions, demonstrating a different cultural-specific representation of the basic emotions. It is worth noting that these findings are approached from a psychological viewpoint, thereby the authors did not consider the differences that can be found in the facial representation of expressions and the effect that these differences may have in automatic FER systems.

On the other hand, from the HCI viewpoint, Da Silva and Pedrini [20] proposed an analysis of recognition performance when a restricted out-group scenario occurs, training different FER systems with one specific culture and testing with a set of different culture. Using only occidental and oriental databases, the

authors concluded that multicultural training should be considered when an efficient recognition performance is needed. Additionally, they pointed out that the six basic expressions are universal with subtle differences which could be influenced by lighting changes or other image problems. Ali *et al.* [13] performed a similar study, where ensemble classifier construction was intended to find how those will be trained to accurately classify multicultural databases (from Caucasian and East-Asian people). The authors concluded that promising results are obtained when multicultural databases are used for training, increasing the accuracy achieved by the cross-cultural database experiments without regard of decreasing the accuracy in the same culture-specific database experiments. However, the misrecognition problems were attributed to factors such as differences in the number of samples per expression and visual representation. In summary, many FER studies have included cross-database tests [11] [12] [13] [14] [20], but when considerable differences of accuracy among multicultural and out-group tests are found, they attribute those problems to external factors such as algorithm robustness or image quality rather than question the universality of facial expressions itself.

Based on the literature of cross-cultural studies, we can summarize some important points which must be considered in order to develop a robust cross-cultural analysis:

- Consistency in the size of the datasets and in the number of samples per expression.
- A variety of geographic region, culture, ethnicity and race of the participants.
- Robust facial feature extraction method (especially to illumination changes).
- Review of in-group, out-group, multicultural and out-group multicultural cross-classification.
- Consideration of static structure of individuals faces (neutral face treatment).
- Visual representation of results and human study validation.
- Independent facial region treatment per expression.

After explaining related works and considering all of the mentioned points, we can proceed to explain our proposed analysis.

3. Proposed Feature Extraction Method

The proposed method is divided into four phases: appearance features extraction, geometric features extraction, hybrid vector conformation and the consideration of the static structure of faces. The basis of this proposal is the Discrete Fourier Transform (DFT), which has been successfully applied for facial recognition and FER, where three different Fourier feature domains were fused for face recognition [15] and phase spectrums of non-overlapped sub-blocks for FER [17]. However, those approaches were applied using only appearance features. Our proposal instead attempts to use a combination of geometric and appearance Fourier features. In addition, in order to locally analyze the changes that may appear on the face while showing facial expressions, our proposal is based on four specific facial regions: forehead, eyes-eyebrows, mouth and nose.

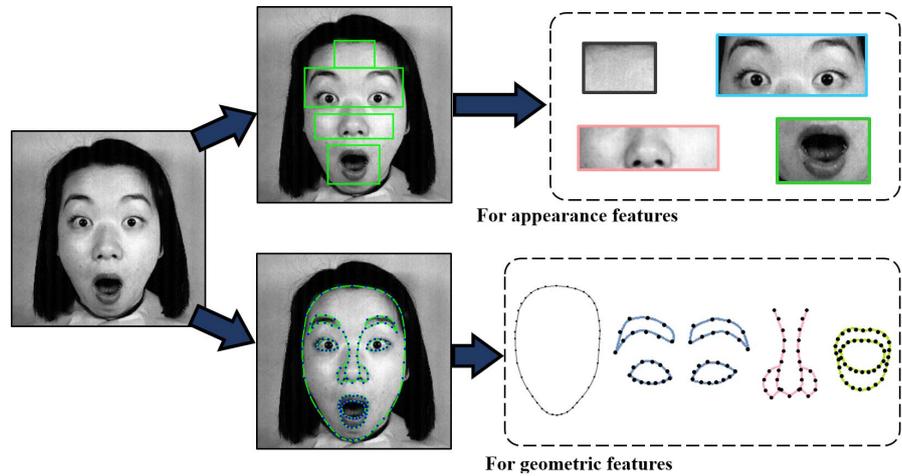


Figure 1. Example of a facial region segmentation based on appearance and geometric features.

In this way, we can analyze the culture-specific capability to recognize the basic expressions by using only one facial region. Furthermore, this analysis is extended to the combinations that can be formed by the four independent regions, thus we can find the combination that achieves better recognition performance based on cross-cultural tests. An example of the region segmentation of both feature-based methods is shown in **Figure 1**.

3.1. Appearance Features

As mentioned before, the appearance features are individually obtained from four independent facial regions. Thus, for each facial region we applied the proposed feature extraction method independently. Therefore, the first step of this process is facial region segmentation which is based on the distance between irises and its relation with the four facial regions [10]. **Figure 1** illustrates an example of this automatic segmentation.

Our proposal is based on the sub-block analysis of eigenphases algorithm presented in [16] which concludes that for face recognition, the ideal size of a sub-block for facing illumination problems is the smallest possible. Since the minimum size of a sub-block is equal to 2×2 pixels, and due to the complex part of the Fourier transform being equal to zero in this particular case, we propose to use Fourier coefficients instead of the phase spectrum for the feature extraction process. For instance, consider I_{FR} as the FR-th facial region image of size $M \times N$, and for convenience I represents one of the four facial regions which is subdivided into sub-blocks of size $L \times L$. Then, the 2-D DFT of each sub-block of the facial region is given by

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

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

$$f_{p,q}(u,v) = \text{Re}(u,v) + j \times 0 \times \text{Im}(u,v), \quad (2)$$

where $\text{Re}(u,v)$ and $\text{Im}(u,v)$ are the real and imaginary components of $f_{p,q}(u,v)$ respectively. Thus, the final Fourier coefficient matrix is given by

$$F = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,N/L} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,N/L} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M/L,1} & f_{M/L,2} & \cdots & f_{M/L,N/L} \end{bmatrix}, \quad (3)$$

where F is the Fourier coefficient matrix of the facial region, which has the same dimensions as I . Finally, in order to have correlated information with the set of training images and for dimensionality reduction we apply PCA. To this end, this confusion matrix should be converted into a column vector, so that

$$F_{Ap} = \text{vec}(F(m,n)), \quad (4)$$

where F_{Ap} is the column vector and $0 \leq m, n < M, N$. Moreover, Fourier coefficient vectors F_{Ap} are independently projected into the discriminative facial region subspace obtained from the training set. These projections are considered as appearance feature vectors of the current facial region, given by

$$Y_{Ap} = W_{Ap}^T (F_{Ap} - \mu_{Ap}), \quad (5)$$

where W_{Ap} is the facial region subspace and μ_{Ap} is the mean vector of all training images. It is worth noting that this PCA process is applied retaining 90% of the variance, so W_{Ap} is calculated using the necessary eigenvectors for retaining that percent of the variance.

3.2. Geometric Features

Geometric features are also focused in four independent regions of the face. However, they define the shape of specific facial regions instead of texture information. The shapes of eyes-eyebrows, nose, lips and the face outline were defined by manually picking up 163 feature points [21]. **Figure 1** shows an example of these four shapes. It is worth noting that this paper intends to precisely analyze facial expression from two different racial groups which may imply subtle differences among specific facial regions. In this way, feature point extraction is crucial for the accuracy of feature extraction settling the analysis efficiency. Therefore, the feature points were extracted manually, even though we have also developed an automatic extraction method of a large number of feature points. The results of this automatic process tend to be affected by shooting conditions and it is still not easy to guarantee the sufficient accuracy for the analysis.

As well as the previous method, our proposal for geometric features is based on the Fourier transform. However, the application of DFT in this particular case is known as Fourier Descriptor (FD), which is a contour-based shape descriptor widely used for content-based image retrieval (CBIR) [22]. Moreover, FD has been used a few times for face recognition [23], and as far as the authors know, it has not been applied to FER. For that reason, we proposed to use a var-

iation of FD and PCA as a feature extraction method for these kinds of features. 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, where $K = 54, 42, 29$ and 38 for eyes-eyebrows, nose, lips and the face outline, respectively. Thus, we assume that a specific shape of the FR-th facial region is represented as the sequence of coordinates:

$$s_{FR}(k) = [x_{FR}(k), y_{FR}(k)], \quad (6)$$

where $k = 0, 1, 2, \dots, K-1$. From one of the four facial regions, complex numbers have to be generated from each coordinate pairs, as in

$$s(k) = [x(k) - x_c] + j[y(k) - y_c], \quad (7)$$

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

$$x_c = \frac{1}{K} \sum_{k=0}^{K-1} x(k), \quad y_c = \frac{1}{K} \sum_{k=0}^{K-1} y(k). \quad (8)$$

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

$$F_{Ge}(u) = \sum_{k=0}^{K-1} s(k) e^{-2j\pi uk/K}, \quad (9)$$

for $u = 0, 1, 2, \dots, K-1$, where $F_{Ge}(u)$ represents the Fourier coefficient vector which has to be projected into the current facial region subspace obtained by applying PCA to the training set. Therefore, geometric feature vectors are defined as

$$Y_{Ge} = W_{Ge}^T (F_{Ge} - \mu_{Ge}), \quad (10)$$

where W_{Ge} is the current facial region subspace and μ_{Ge} is the mean vector of all training images. It is important to mention that unlike the appearance-based process given by Equation (5), here 99% of the variance is retained.

3.3. Hybrid Method

The combination of geometric and appearance features has been successfully applied for FER [7] [8] [9]. Some approaches perform the fusion at classification level but better results are obtained when the combination is done at the feature extraction phase [8]. From the literature we can see that feature extraction methods based on hybrid features are focused on appearance features and improve the final feature vector by adding facial landmarks. In this way, the characteristics of both types of features are different and the final feature vector could be directly affected by the individual problems of one kind of features. Therefore, the fusion process of our proposal is based on the application of PCA for correlating the information of both types of features. In addition, the proposed feature extraction method employs the same principle based on the Discrete Fourier Transform. Thus, the appearance features are extracted by using the 2D DFT and the geometric features by FD. It is worth noting that in order to efficiently correlate the features, PCA process has to be applied individually before the fusion.

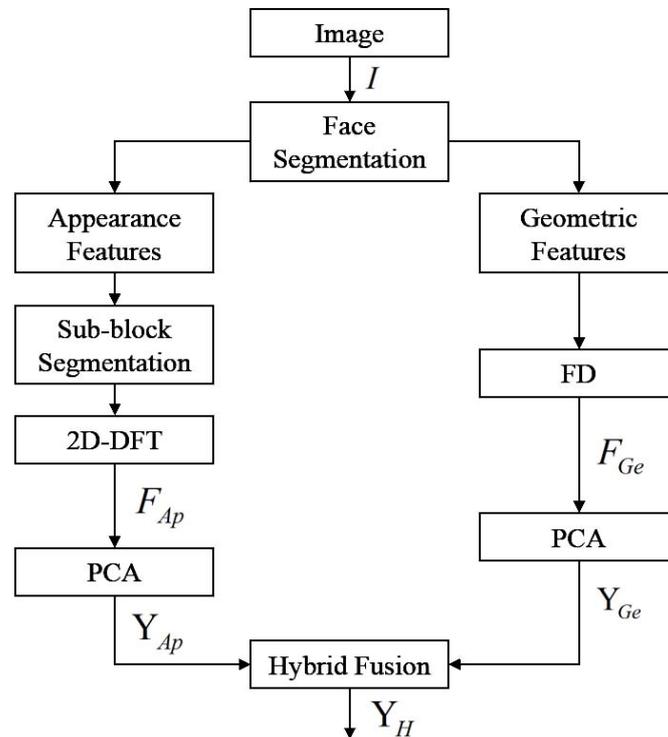


Figure 2. Framework of the fusion process of hybrid Fourier features.

The process of the complete fusion process is shown in **Figure 2**, where first we obtained individual feature vectors of appearance and geometric features, as defined in Equations ((5) and (10)), respectively. Thus, the process begins with the concatenation of both feature vectors, so that

$$H = [Y_{Ge}^T, Y_{Ap}^T]^T. \quad (11)$$

Finally, the projection of H onto a new subspace of hybrid features of the current facial region represents the hybrid feature vector, which is given by

$$Y_H = W_H^T (H - \mu_H), \quad (12)$$

where W_H is the current facial region subspace obtained by retaining the 99% of the variance, and μ_H is the mean hybrid vector of all training images. It is worth noting that Y_H only represents the feature vector of an individual facial region. Therefore, the final feature vector which is based on all possible combinations of the four facial regions is defined as

$$Y = \bigcup_{l=1}^{FR} Y_H(l), \quad FR = 1, 2, 3, 4, \quad (13)$$

where Y_H represents one of the four hybrid feature vectors based on a specific facial region and Y the concatenation of them. Thus, it can be conformed up to four individual facial regions.

3.4. Static Structure of Individual Faces

In order to design a FER system robust to the static structure of individual faces, some works assumed that facial expressions can be represented as a linear com-

bination based on the difference between expressive facial images and neutral image of the same subject [23] [24]. It is well known that the structural characteristics and texture information which define a specific expression appear when the face images change from neutral to expressive, hence the difference image may represent those changes and it can reduce the dependency on the subject's identity as well. Therefore, difference images may also reduce the physical differences among faces from different races and be focused only in the way to constitute the facial expressions. For that reason, we propose to make that difference between feature vectors instead of the raw images, so that

$$Z(l) = Y_{Ex}(l) - Y_{Ne}(l), \quad (14)$$

for $l = 0, 1, 2, \dots, Q-1$, where Q is the total number of expressive images in the dataset, Y_{Ex} and Y_{Ne} represents the final feature vectors of expressive and neutral facial image, and Z the difference feature vector. It is important to mention that for this process, information of neutral images must be considered on each facial region subspace (W) defined in Equations ((5), (10) and (12)).

As a visual example of the effect of mentioned subtraction, **Figure 3** shows two difference images of subjects from different races showing the same facial expression. As we can observe, the expressive images in the first column present physical changes related to gender and race. These noticeable differences between images of the same expression may affect the classification process. On the other hand, the difference images (third column) present clear similarities based on the facial actions produced by the expression rather than physical differences.

4. Proposed Visual Analysis

In order to make a robust analysis of facial expressions, visual changes presented in the face must be taken into account. Therefore, we present two methods for

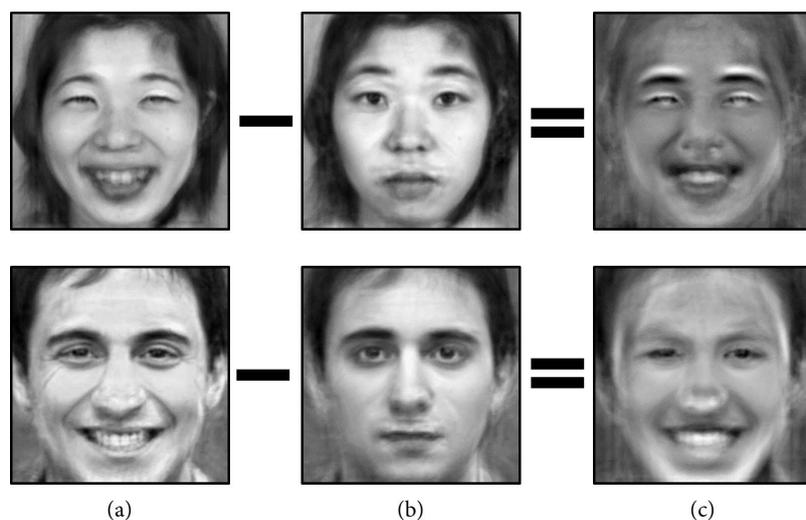


Figure 3. Example of two difference images reconstructed from its feature vectors. (a) Expressive images from Y_{Ex} ; (b) Neutral images from Y_{Ne} ; (c) difference images from Z .

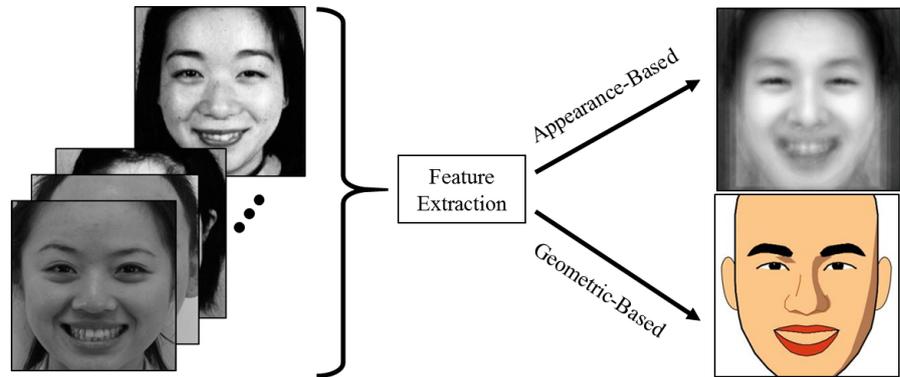


Figure 4. Average projected image of happiness from ASN dataset represented by using appearance and geometric features.

visually analyzing the differences which may appear among both racial groups. Those methods are based on the type of features employed for the analysis: appearance- and geometric-based. **Figure 4** shows an example of the visual representation of the average happy expression from both feature-based approaches.

4.1. Appearance-Based

The visual representation of appearance features is based on the well-known algorithm of Eigenfaces. The ability of this method for reconstructing images from projected vectors of a previously defined Eigenspace has been successfully applied for analyzing facial expressions [25]. Therefore, using reconstructed images from feature vectors gives the opportunity to analyze the differences and similarities which may appear among the basic expressions of different cultures in detail. To this end, we have to obtain average projected vectors by each expression, which can be calculated from a cultural-specific or multicultural dataset. These average vectors are given by:

$$Z_{\mu}(r) = \frac{1}{P(r)} \sum_{i=0}^{P(r)} Z(i), \quad (15)$$

for $r = 0, 1, 2, \dots, 5$ which represents the number of basic expressions and $P(r)$ the number of images per expression, so that $Q = P \times r$. As mentioned before, in order to have a better analysis of facial expressions the number of frames per expression should be equal, then $P(r) = Q/6$. Finally, reconstructed images are the reshaped matrix of reconstructed average projected vectors given by:

$$R(r) = W Z_{\mu}(r) + \mu, \quad (16)$$

for $r = 0, 1, 2, \dots, 5$, where W is the subspace where the Z projections were made, and μ is the mean feature vector of all training images. An example of one reconstructed image from an average happy expression is illustrated in **Figure 4**.

4.2. Geometric-Based

It is known that the raw data of geometric-based methods are easier to visualize than those of appearance-based. However, their projections made by PCA are more difficult to be visually analyzed. Therefore, in order to accurately analyze

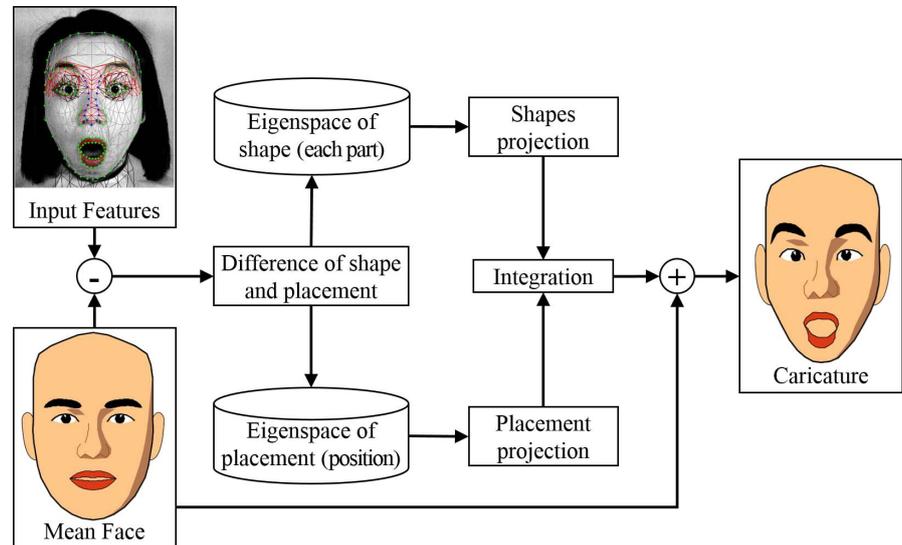


Figure 5. Framework of DrawFace tool [26] for developing facial caricatures.

the geometric features, we employed the DrawFace tool [26], which has been successfully applied for representing facial expressions from geometric feature vectors using PCA [19]. DrawFace tool draws caricatures based on the Eigenfaces process, hence this tool requires to input individual eigenspaces of facial regions and the mean facial shape as initialization setup.

Figure 5 presents the general framework of DrawFace tool. Similar to the PCA process, the mean face of the complete dataset has to be subtracted from the input, but the eigenspaces are calculated individually for each facial region and the placement of them. Thus, the final caricature is drawn by integrating the projections of the input features into its respective eigenspace. In this way, these caricatures can be considered as a result of a projected feature vector into a subspace made by a set of specific facial shapes. For this analysis, we use the eigenspaces and average projected vectors as obtained by Equations ((10) and (15)). Examples of caricatures of an average happy expression and individual surprise face from ASN dataset can be seen in **Figure 4** and **Figure 5** respectively.

5. Human Study

As a baseline for the experimental results, we present a cross-cultural human study based on a survey applied to subjects from a different race and culture. Forced-choice facial expression classification from each participant was collected by using the same datasets employed for the proposed experiments as stimuli. Relevant information about this study is presented as follows.

5.1. Participants

The experiment was applied to 40 students of the University of Electro-Communications in Tokyo Japan. Participants were divided into two groups: 20 East-Asians that include Japanese, Taiwanese and Chinese students (50% males); and 20 Western-Caucasians that include German, Swedish, American and Mex-

ican students (50% males). Their ages ranged from 20 to 26 years old (mean 22). It is important to mention that the Western-Caucasians are currently exchange students and had newly arrived in an Asian country for the first time with a residence time no longer than 3 months on average.

5.2. Procedure

A GUI-based script for collecting and presenting the survey was developed in MATLAB. After a brief explanation of the experiment by the software, the stimuli were automatically presented one by one. The instructions of the experiment were presented in Japanese and Chinese for East-Asian participants and in English for Westerns. Each stimulus appeared in the central visual field and remained visible just for 3 seconds, followed by a 6-way forced choice decision question related to the 6 basic expressions. The question presented was “What is the expression of the face?” and the participant has to choose one answer before clicking the button “Next” for the following stimulus. We randomized trials within each participant and all of them have to recognize the 240 stimuli of the dataset which have been presented by groups of 30 images with breaks of 30 seconds between them. **Figure 6** presents two screenshots of the application software used for this study.

6. Experimental Results

6.1. Datasets

FER test, visual analysis and human study were evaluated using the same datasets. A total of 656 images were used for this paper. 480 expressive and 176 neutral facial images were selected from four standard datasets which were divided into two racial groups, representing three possible training sets: Western-Caucasian dataset (WSN), East-Asian (ASN) dataset and the multicultural dataset (MUL).

WSN dataset is comprised of 330 facial images from 90 different subjects (40% males) of the extended Cohn-Kanade dataset (CK+) [27], which in turn is com-

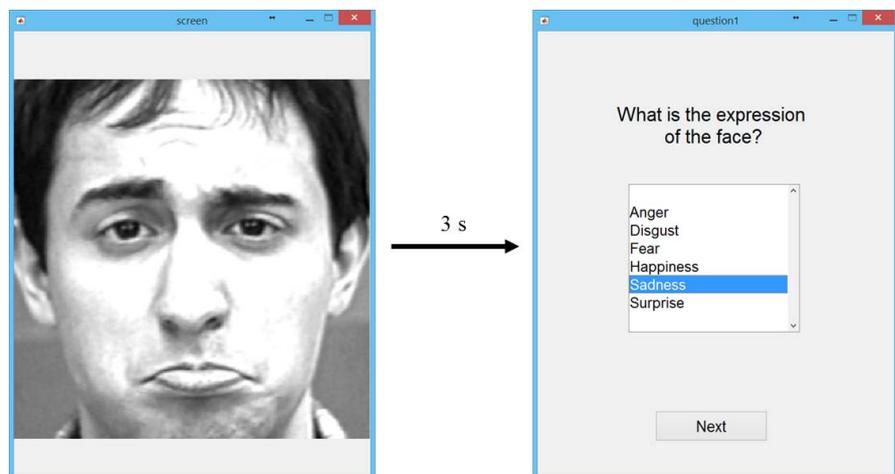


Figure 6. Screenshots of the software application used for the human study.

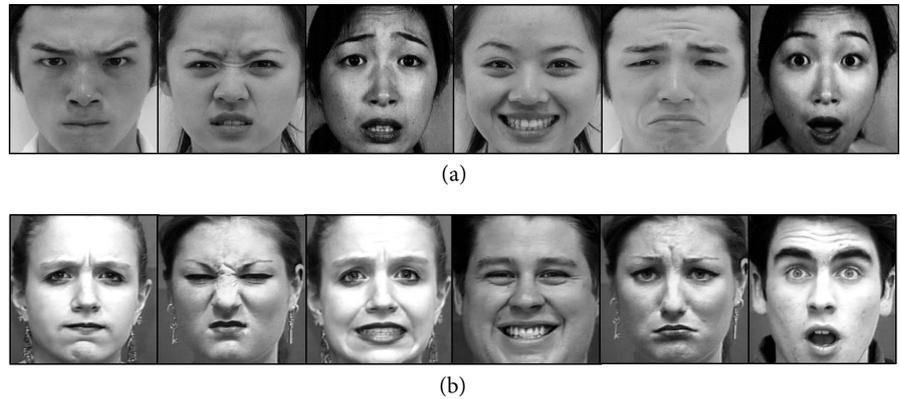


Figure 7. Example of images included in each dataset. For each row from left to right: anger, disgust, fear, happiness, sadness and surprise. (a) ASN dataset and (b) WSN dataset.

posed of 327 facial image sequences from 123 subjects performing the 6 basic emotions plus the neutral state. For this paper, the subset selected from CK+ only includes images of 90 Euro-American subjects.

ASN dataset contains 326 images selected from 86 subjects (47% males) of three different datasets: Japanese Female Facial Expression (JAFFE) dataset [28], which comprises 213 images of 10 Japanese female models; Japanese and Caucasian Facial Expression of Emotion (JACFEE) dataset [29], which contains 56 images from different individuals including 28 Japanese and 28 Caucasian subjects; and Taiwanese Facial Expression Image Database (TFEID) [30], which includes 336 images from 40 Taiwanese models. In this way, ASN dataset is formed by 86 subjects from Japan and Taiwan.

Finally, MUL dataset is a combination of WSN and ASN. It is worth noting that the number of images per facial expression is equitable among each dataset, being 40 images per expression for ASN and WSN so that 240 expressive images correspond to each dataset. However, the neutral facial images vary from each dataset. **Figure 7** shows an example of the six basic expressions displayed by subjects from WSN and ASN datasets.

All facial images were pre-processed in order to have the same inter-ocular distance and eye position as well as cropped with 280×280 pixels. For the regions of forehead, eyes-eyebrows, mouth and nose their sizes were normalized at 100×70 , 200×80 , 140×80 and 175×50 pixels respectively.

6.2. Cross-Cultural Classification

Feature vectors of the six basic expressions were classified by multi-class SVMs with RBF kernels [31], and evaluated by leave one subject out cross-validation. Average recognition rates and confusion matrices are presented to show the accuracy of FER. The visual analysis is focused on the changes of facial images based on the FACS, for further information refer to [3]. **Table 1** shows all results obtained by the proposed method and human study.

It is worth noting that the results of **Table 1** were obtained by using hybrid features from facial regions obtained by combining appearance and geometric

Table 1. Classification accuracy of individual facial regions and its combinations.

| Training Set: | MUL | WSN | MUL | ASN | ASN | MUL | WSN |
|-------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Test Set: | MUL | WSN | WSN | WSN | ASN | ASN | ASN |
| Human Study | - | 76.8 | - | 66.7 | 71.7 | - | 67.2 |
| Forehead (F) | 62.5 | 71.7 | 64.2 | 45.0 | 62.9 | 60.8 | 48.8 |
| Eyes-Eyebrows (E) | 68.5 | 75.0 | 68.3 | 54.2 | 71.3 | 68.8 | 50.4 |
| Nose (N) | 72.1 | 81.7 | 80.0 | 46.7 | 63.8 | 64.2 | 60.8 |
| Mouth (M) | 89.4 | 95.4 | 94.6 | 68.3 | 84.2 | 84.2 | 78.8 |
| F-E | 78.1 | 78.8 | 76.3 | 54.2 | 80.0 | 80.0 | 58.8 |
| F-N | 81.9 | 84.6 | 87.1 | 57.1 | 75.8 | 76.7 | 61.3 |
| F-M | 91.0 | 92.5 | 95.4 | 64.6 | 88.3 | 86.7 | 76.3 |
| E-N | 80.2 | 81.7 | 82.9 | 54.6 | 80.4 | 77.5 | 55.8 |
| E-M | 94.4 | 98.3 | 95.8 | 72.5 | 94.6 | 92.9 | 82.5 |
| N-M | 92.1 | 94.6 | 94.6 | 68.3 | 86.3 | 89.6 | 81.7 |
| E-N-M | 95.0 | 98.8 | 95.4 | 72.1 | 95.0 | 94.6 | 85.0 |
| F-N-M | 93.8 | 95.0 | 96.8 | 62.9 | 88.3 | 91.7 | 75.4 |
| F-E-M | 93.8 | 96.3 | 93.8 | 69.2 | 93.8 | 93.8 | 81.3 |
| F-E-N | 83.8 | 87.9 | 85.8 | 57.9 | 84.6 | 81.7 | 62.9 |
| F-E-N-M | 95.2 | 97.5 | 96.8 | 69.2 | 94.6 | 94.6 | 84.2 |

features. Because the facial region of the forehead cannot be represented by geometric features, for convenience, this region was fused with the outline shape of the face. Other regions were fused with the same facial region of its counterpart kind of features.

From **Table 1** we can observe a trend of the results which follows the order from top to bottom the accuracy obtained by the tests of in-group, out-multicultural, multicultural and out-group. Thus, the best result is reached by the in-group test of WSN (WSN vs WSN), followed by the out-multicultural test of the same dataset (MUL vs WSN) and the third place is for the multicultural test (MUL vs MUL). This trend is also followed by the human study, showing that the WSN dataset is easier for classifying the prototypic expressions.

The analysis per facial region combinations shows that the combination of E-N-M performs better in most of the training combinations. On the other hand, the best single region for FER is the mouth, followed by the nose. However, the eyes regions present interesting results, especially for the out-group and out-multicultural analysis, where this region seems to define the expressions of ASN better. In order to analyze the performance per expression of those results, **Table 2** and **Table 3** show the accuracy of the mouth region and E-N-M combination. In these tables, we can notice that the mentioned trend is not followed by all the expressions. For example, the out-multicultural test of the mouth shows better accuracy when anger is tested using ASN dataset rather than WSN. In summary, we can observe that the clearer difference between cultures is presented in the out-group test.

Table 2. Classification accuracy per expression using the region of Mouth.

| Training Set: | MUL | WSN | MUL | ASN | ASN | MUL | WSN |
|---------------|------|-------|------|------|------|------|------|
| Test Set: | MUL | WSN | WSN | WSN | ASN | ASN | ASN |
| ANG | 92.5 | 95.0 | 95.0 | 72.5 | 90.0 | 90.0 | 85.0 |
| DIS | 81.3 | 90.0 | 90.0 | 60.0 | 72.5 | 72.5 | 32.5 |
| FEA | 87.5 | 95.0 | 97.5 | 72.5 | 72.5 | 77.5 | 80.0 |
| HAP | 88.8 | 95.0 | 95.0 | 92.5 | 85.0 | 82.5 | 70.0 |
| SAD | 88.8 | 100.0 | 92.5 | 85.0 | 87.5 | 85.0 | 60.0 |
| SUR | 97.5 | 97.5 | 97.5 | 90.0 | 97.5 | 97.5 | 82.5 |
| Avg. | 89.4 | 95.4 | 94.6 | 78.8 | 84.2 | 84.2 | 68.3 |

Table 3. Classification accuracy per expression using the combination of E-N-M.

| Training Set: | MUL | WSN | MUL | ASN | ASN | MUL | WSN |
|---------------|-------|-------|-------|-------|------|-------|------|
| Test Set: | MUL | WSN | WSN | WSN | ASN | ASN | ASN |
| ANG | 93.8 | 100.0 | 95.0 | 97.5 | 92.5 | 92.5 | 82.5 |
| DIS | 92.5 | 97.5 | 92.5 | 87.5 | 92.5 | 92.5 | 40.0 |
| FEA | 93.8 | 95.0 | 90.0 | 50.0 | 92.5 | 97.5 | 72.5 |
| HAP | 92.5 | 100.0 | 100.0 | 100.0 | 97.5 | 85.0 | 60.0 |
| SAD | 97.5 | 100.0 | 95.0 | 80.0 | 97.5 | 100.0 | 90.0 |
| SUR | 100.0 | 100.0 | 100.0 | 95.0 | 97.5 | 100.0 | 87.5 |
| Avg. | 95.0 | 98.8 | 95.4 | 85.0 | 95.0 | 94.6 | 72.1 |

6.3. Racial Comparisons

In order to have a robust results evaluation, an analysis of the differences between both racial groups based on its expressive faces must be done. Therefore, **Figure 8** shows the visual representation of the average expressions of both datasets obtained from appearance and geometric features. The expressions represented by each row from left to right are: anger, disgust, fear, happiness, sadness and surprise. From this illustration, we can observe the same distinctive patterns of difference from specific expressions. For instance, disgust and fear expressions look different. Disgusted faces from WSN fulfill the necessary AUs to be classified as disgust (AU9, AU15, AU16). However, the same average face from ASN presents an extra AU22 and AU23 which are known to appear in anger expression. The average face of fear from WSN again covers the EMFACS for fear expression (AU1, AUN2, AU4, AU5, AU7, AU20, AU26). However, that of ASN lacks AU4 and AU20, given the impression of surprised in the eyes region and the mouth is not well defined.

Another way to analyze the distinctive properties of average projected vectors is simply by plotting them. In this way, we can visualize the behavior of each feature vector and the capability of discrimination of the six basic expressions based on them. **Figure 9** shows the six average expression vectors of each dataset. It is easy to see that the six vectors of WSN present better distinctiveness

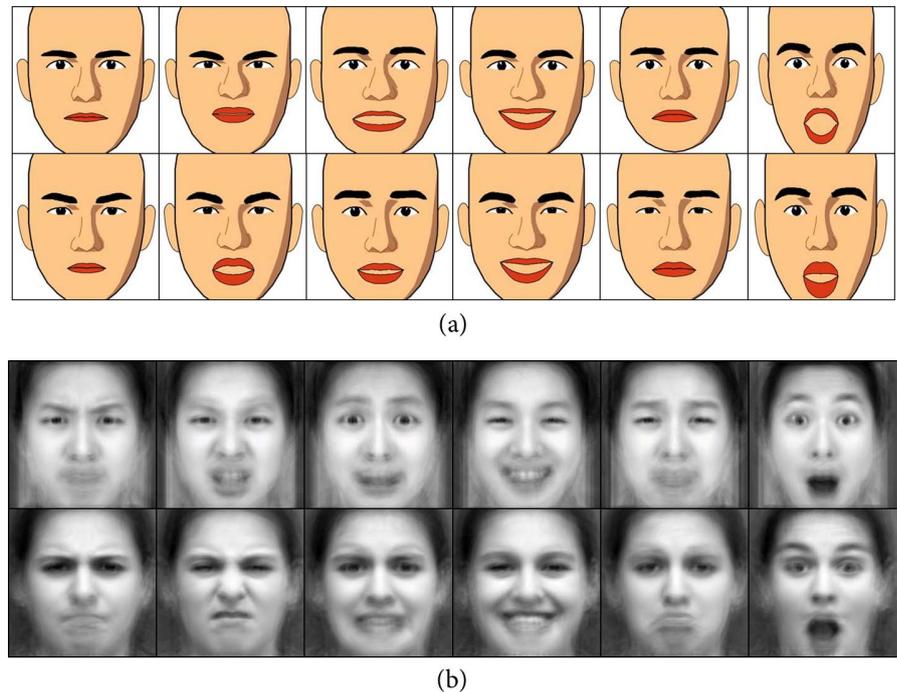


Figure 8. Visual representation of average expressions from ASN (upper row) and WSN (lower row) datasets using (a) appearance features and (b) geometric features. For each row from left to right: anger, disgust, fear, happiness, sadness and surprise.

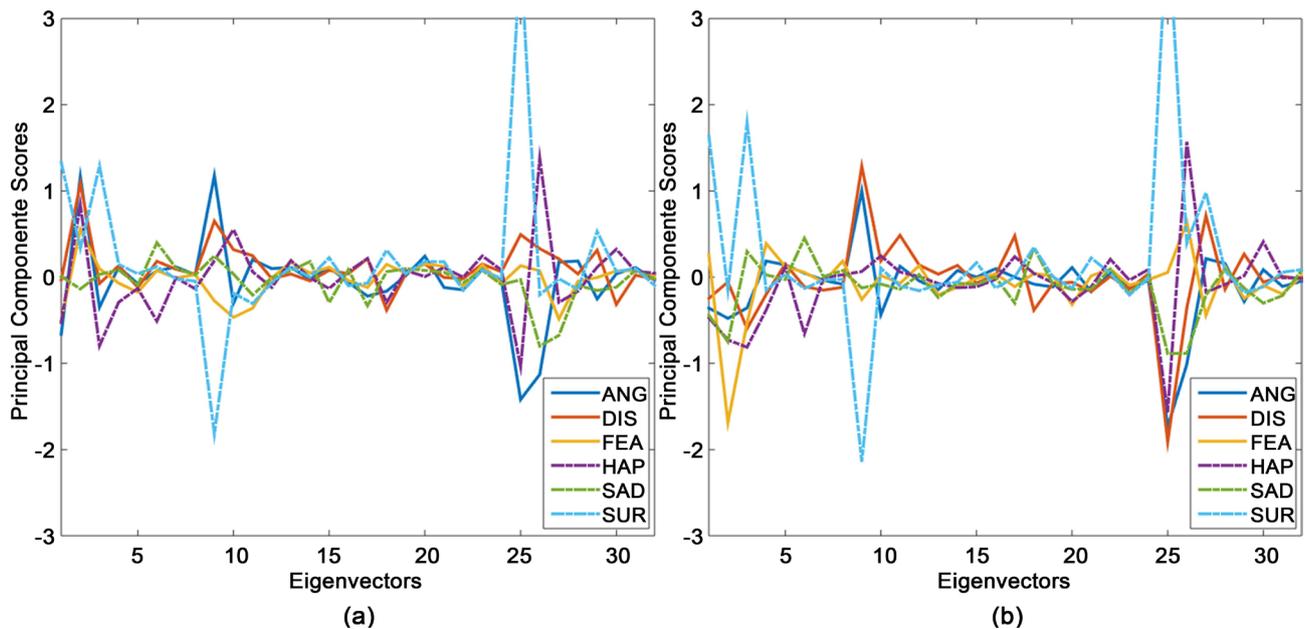


Figure 9. Average projected vectors of six expressions, (a) from ASN and (b) from WSN.

among themselves rather than those of ASN, which have problems especially for the expressions of fear, disgust and anger.

Thanks to the visual analysis we can better analyze the misrecognition problems which appear among the six basic expressions. **Figure 10** and **Figure 11** show the confusion matrices for WSN and ASN of Mouth and Eyes region, respectively. Note that these matrices present a gray scale color map for simplify-

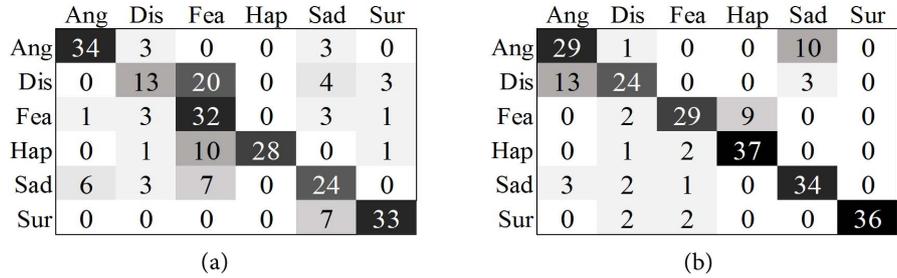


Figure 10. Confusion matrices of mouth region for out-group test. (a) ASN vs WSN (b) WSN vs ASN.

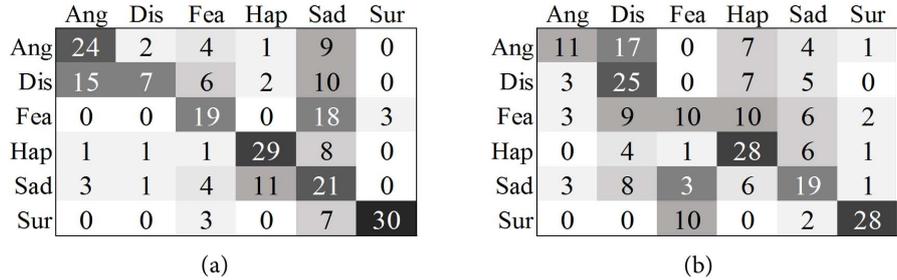


Figure 11. Confusion matrices of eyes region for out-group test. (a) ASN vs WSN (b) WSN vs ASN.

ing the visualization, showing higher results in black and lower in white. The confusion matrices represent the true expression on the vertical axis and the decision made by the classifier is presented on the horizontal axis so that each row of the matrix indicates the level of confusion of each expression with its counterparts.

From these confusion matrices, we can observe that the misclassification problems mainly reside in the expressions of fear and disgust for both facial regions. Some examples of the faces misrecognized even for the human study are presented in **Figure 12**, which shows the original expressive facial image, its feature vector extracted using E-M-N combination and the visual representation using appearance features (difference image).

Finally, **Table 4** presents a comparison of our results with those found in the literature. A few studies have analyzed the cross-cultural recognition capabilities of their proposals. Nevertheless, none of them have covered the out-group multicultural analysis. Even though the recognition accuracy is vastly ranged, the same trend is found: the results of in-group tests are better than those of out-group and WSN datasets achieve higher accuracy than ASN.

It is important to mention that the results from **Table 4** widely range from ours, mainly because of some specific reasons: the inconsistency on the number of samples per expression for training (some expressions used more than 200% of samples than others); the algorithms used for feature extraction (all of them only used appearance features for this task); and the null consideration of the static structure of individual faces for the analysis (some of the studies consider the neutral face as an extra class for the recognition task but not for the analysis itself). As an extra reference from the setup of mentioned works, **Table 5** shows the datasets used for defining the datasets of WSN and ASN.

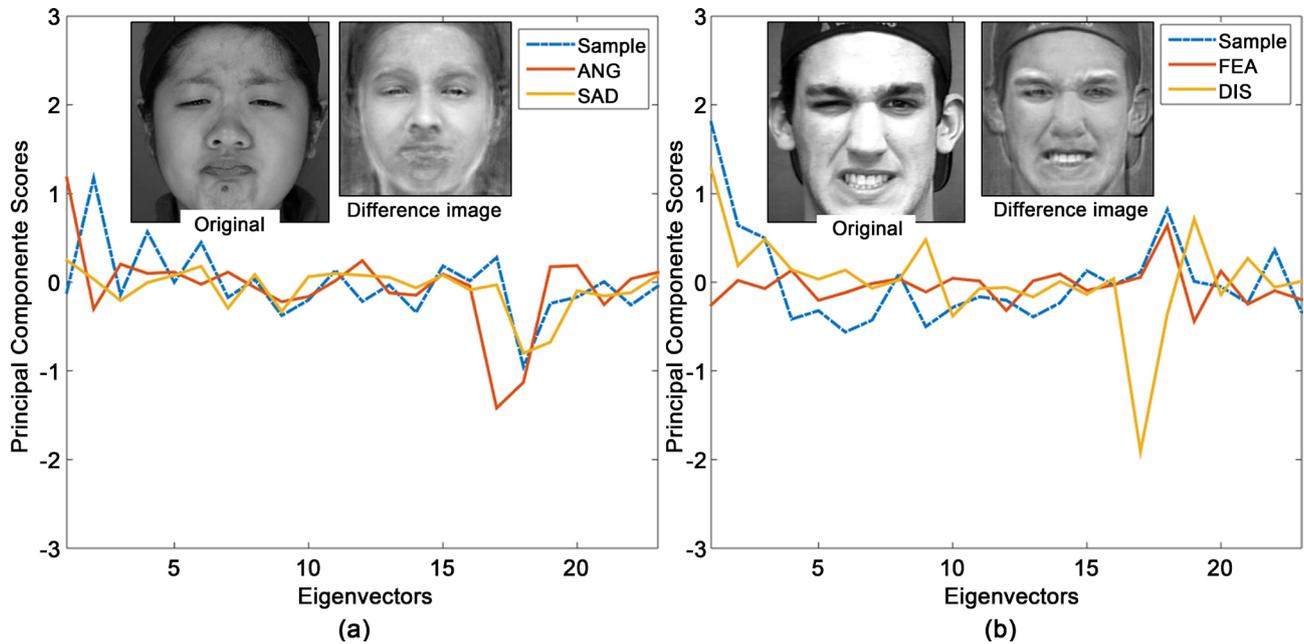


Figure 12. Examples of misclassification. (a) sample from ASN dataset showing anger misrecognized with sadness. (b) sample from WSN dataset showing fear misrecognized with disgust.

Table 4. Classification accuracy of the proposed method and previous works with cross-cultural tests.

| Training Set: | ASN | WSN | ASN | WSN | MUL | MUL | MUL |
|----------------|------|------|------|------|------|------|------|
| Test Set: | ASN | WSN | WSN | ASN | MUL | ASN | WSN |
| Gabor [11] | 89.7 | 91.5 | 54.1 | 55.9 | - | - | - |
| LBP + SVM [12] | 81.0 | 91.4 | - | 41.3 | - | - | - |
| HOG + NNE [13] | - | - | 55.9 | 63.6 | 93.8 | - | - |
| CNN [14] | 86.7 | 98.9 | - | 79.6 | - | - | - |
| HOG + SVM [20] | 88.6 | 93.5 | 39.9 | 42.3 | 84.7 | - | - |
| Human Study | 71.7 | 76.8 | 66.7 | 67.2 | - | - | - |
| Proposed | 95.0 | 98.8 | 72.5 | 85.0 | 95.2 | 94.6 | 96.8 |

Table 5. Datasets included in each cultural group of previous cross-cultural studies.

| Dataset: | ASN | WSN |
|----------------|----------------------|----------|
| Gabor [11] | JAFFE | CK |
| LBP + SVM [12] | JAFFE | CK |
| HOG + NNE [13] | JAFFE, TFEID | RAFD |
| CNN [14] | JAFFE | CK+ |
| HOG + SVM [20] | JAFFE | CK+, MUG |
| Proposed | JAFFE, JACFEE, TFEID | CK+ |

7. Conclusions and Future Work

In this paper, we presented a methodical analysis of Western-Caucasian and

East-Asian basic expressions focused on four facial regions. Based on the literature and from a psychological viewpoint, we know that there exist in-group advantages for recognizing facial expressions when using cross-cultural datasets and Western-Caucasian facial expressions are closer related to the six basic emotions than those of East-Asian. Moreover, from an HCI viewpoint we know that multicultural training is necessary for increasing the accuracy of FER systems. Following these findings, the cross-cultural tests analyzed in this paper present the same trend about the design of the training phase which suggests that WSN basic expressions are easier to recognize than those of ASN. In addition, as indicated by previous works, we found strong cross-cultural similarities on the expressions of happiness and surprise. On the other hand, even though the system trained by using the multicultural dataset (MUL) reached higher accuracy, the proposed analysis of out-group multicultural test follows the same trend by exposing difficulties for recognizing ASN expressions. This issue points out the cultural-specific differences that can be found on showing the six basic expressions. To this end, and thanks to the facial region segmentation presented in this paper, we fully analyzed the cross-cultural recognition performance of individual facial regions and its combinations. In this way, our analysis also contributes to identifying the differences between WSN and ASN expressive faces which are primarily found in the regions of mouth and eyes-eyebrows, specifically for expressions of disgust and fear.

In summary, we can conclude that when working with multicultural datasets for FER, it is better to set up specific training for each cultural group. This issue could be pointed out by including an extra pre-processing stage for race recognition. In addition, the cultural-specific training strongly depends on the application of the FER system due to the cross-cultural similarities found on the so-called positive expressions. On the other hand, we frankly recognize that this analysis deals with the problem of reliability of the datasets. Most of the standard datasets are taken under controlled environments and expressions are shown by professional actors which sometimes exaggerate and break the spontaneity of a true facial expression. However, despite the limitations of the datasets, the proposed analysis helps to find cultural differences of specific facial expressions and introduces a methodical process for analyzing new FER methods.

As a future work, we plan to expand the limited size of the datasets, as well as to include datasets without controlled environments and with subjects of different nationalities from similar cultures. In addition, we would like to analyze the reliability of the set of basic expressions for ASN, from an HCI viewpoint.

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. Last but not least we thank the students who participated in the human study.

References

- [1] Darwin, C. (1872) The Expression of the Emotions in Man and Animals. Oxford University Press, Oxford. <https://doi.org/10.1037/10001-000>
- [2] Ekman, P. (1972) Universal and Cultural Differences in Facial Expression of Emotion. In: *Proceedings of Symposium on Motivation*, Nebraska University Press, Lincoln, 9-15.
- [3] Ekman, P., Friesen, W.V. and Hager, J.C. (1978) Facial Action Coding System (FACS). A Technique for the Measurement of Facial Action. Consulting, Palo Alto, 22.
- [4] Jack, R.E. (2013) Culture and Facial Expressions of Emotion. *Visual Cognition*, **21**, 1248-1286. <https://doi.org/10.1080/13506285.2013.835367>
- [5] Dailey, M.N., et al. (2010) Evidence and a Computational Explanation of Cultural Differences in Facial Expression Recognition. *Emotion*, **10**, 874-893. <https://doi.org/10.1037/a0020019>
- [6] Jack, R.E., Garrod, O.G., Yu, H., Caldara, R. and Schyns, P.G. (2012) Facial Expressions of Emotion Are Not Culturally Universal. *Proceedings of the National Academy of Sciences*, **109**, 7241-7244. <https://doi.org/10.1073/pnas.1200155109>
- [7] Tian, Y., Kanade, T. and Cohn, J.F. (2011) Facial Expression Recognition. In: *Handbook of Face Recognition*, Springer, London, 487-519. https://doi.org/10.1007/978-0-85729-932-1_19
- [8] 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>
- [9] Deshmukh, S., Patwardhan, M. and Mahajan, A. (2016) Survey on Real-Time Facial Expression Recognition Techniques. *IET Biometrics*, **6**, 155-163.
- [10] 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>
- [11] 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.
- [12] 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.
- [13] Ali, G., Iqbal, M.A. and Choi, T.S. (2016) Boosted NNE Collections for Multicultural Facial Expression Recognition. *Pattern Recognition*, **55**, 14-27.
- [14] 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.
- [15] 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>
- [16] 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.
- [17] 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>
- [18] Turk, M.A. and Pentland, A.P. (1991) Face Recognition Using Eigenfaces. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*

- ognition, Hawaii, 3-6 June 1991, 586-591. <https://doi.org/10.1109/cvpr.1991.139758>
- [19] Benitez-Garcia, G., Nakamura, T. and Kaneko, M. (2017) Analysis of Differences between Western and East-Asian Faces Based on Facial Region Segmentation and PCA for Facial Expression Recognition. In: *AIP Conference Proceedings* 1807, AIP Publishing, Tokyo, Article ID: 020025. <https://doi.org/10.1063/1.4974807>
- [20] 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**, 0230151-0230159. <https://doi.org/10.1117/1.jei.24.2.023015>
- [21] Benitez-Garcia, G., Nakamura, T. and Kaneko, M. (2017) Analysis of In- and Out-Group Differences between Western and East-Asian Facial Expression Recognition. *Proceedings of the 14th IAPR International Conference on Machine Vision Applications (MVA 2017)*, Nagoya, 8-12 May 2017. (In Press)
- [22] El-Bakry, H.M. (2001) Automatic Human Face Recognition Using Modular Neural Networks. *Machine Graphics and Vision*, **10**, 47-73.
- [23] 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>
- [24] 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.
- [25] Calder, A.J., Burton, A.M., Miller, P., Young, A.W. and Akamatsu, S. (2001) A Principal Component Analysis of Facial Expressions. *Vision Research*, **41**, 1179-1208.
- [26] Kaneko, M. (2008) Computerized Facial Caricatures. *Journal of Institute of Image Information and Television Engineers*, **62**, 1938-1943. <https://doi.org/10.3169/itej.62.1938>
- [27] Kanade, T., Cohn, J.F. and Tian, Y. (2000) Comprehensive Database for Facial Expression Analysis. *Proceedings of 4th IEEE International Conference on Automatic Face and Gesture Recognition*, Washington DC, 28-30 March 2000, 46-53. <https://doi.org/10.1109/AFGR.2000.840611>
- [28] Lyons, M., Akamatsu, S., Kamachi, M. and Gyoba, J. (1998) Coding Facial Expressions with Gabor Wavelets. *Proceedings of 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, 14-16 April 1998, 200-205. <https://doi.org/10.1109/AFGR.1998.670949>
- [29] Biehl, M., et al. (1997) Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability Data and Cross-National Differences. *Journal of Nonverbal Behavior*, **21**, 3-21. <https://doi.org/10.1023/A:1024902500935>
- [30] 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.
- [31] Chang, C.C. and Lin, C.J. (2011) LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, **2**, 27. <https://doi.org/10.1145/1961189.1961199>

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jsip@scirp.org