15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Analysis of In- and Out-group Differences between Western and East-Asian Facial Expression Recognition

Gibran Benitez-Garcia Tomoaki Nakamura Masahide Kaneko
Graduate School of Informatics and Engineering, The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan
{gibran, nakamura, kaneko}@radish.ee.uec.ac.jp

Abstract

Recent cross-cultural studies have questioned the cultural universality of facial expressions from a psychological viewpoint. However, the automatic facial expression recognition (FER) systems are based on the assumption that facial expressions are the same for all human beings, excluding the differences that may appear between different races and cultures. Therefore, this paper presents an analysis of culturally specific facial expression recognition focused on Western and East-Asian expressive faces using an automatic FER system based on 3 different feature extraction methods (appearance-, geometric-, and a proposed hybrid-based). Our study is focused on 4 specific facial regions (eyes-eyebrows, mouth, nose and forehead/outline) and it is powered by Principal Component Analysis (PCA) which enables the visual examination of the most relevant differences among the 6 basic expressions from each racial group. Feature extraction methods are evaluated using Support Vector Machines (SVM) and 4 standard databases. In addition, our findings are compared with a cross-cultural human study applied to 40 participants from both racial groups.

1. Introduction

According to Charles Darwin, facial expressions are innate and invariant for human beings and some mammals [1]. With this basis, many psychologists have agreed on the fact that facial expressions are straight linked with the six basic internal emotional states. This proposal defines the prototypic basic expressions of anger, disgust, fear, happiness, sadness and surprise which are recognized across all different races and cultures [2]. However, some cross-cultural studies have questioned and in some degree refuted this assumed cultural universality of facial expressions [3, 4].

On the other hand, from the viewpoint of the human-computer interaction (HCI), the cultural universality of emotions is taken for granted [5]. Therefore, most of the automatic facial expression recognition (FER) systems are based on the assumption that facial expressions are the same for all human beings. Besides some recent approaches reach a highly average recognition rate, none of them are considering the cultural specificity that some subjects could present on their facial expressions. Thus, in order to attain a complex HCI, FER systems have to take into account the differences which may appear between facial expressions from different races and cultures.

In this paper, we present an analysis of culturally specific facial expression recognition based on Western and East-Asian expressive faces using an automatic FER system. The analysis is focused on in- and out-group performance as well as on specific differences presented for certain facial regions on the 6 basic expressions of each racial group using 4 standard databases (480 images). As a baseline, we present a human study applied to 40 subjects (20 Westerns and 20 East-Asians) using the same datasets employed for the FER algorithms as stimulus.

The FER system is conducted by extracting appearance and geometric features from expressive faces which represent pixel intensities and facial shapes of eyes-eyebrows, mouth, nose, forehead and face outline. In order to obtain the most important characteristics of those regions, we utilized Principal Component Analysis (PCA), and for the recognition task of 6 basic expressions we employed Support Vector Machines (SVM). In addition to the appearance- and geometric-based feature extraction methods, we propose a hybrid approach which fuses the features from both methods. As a summary, this paper presents the following contributions: 1) a methodical analysis of in- and out-group differences between Western and East-Asian FER; 2) a hybrid method of appearance and geometric features based on PCA and SVM for FER; 3) an analysis of the differences appeared on the expressive faces of both racial groups.

The rest of the paper is organized as follows. Section 2 discusses the related work from a psychological viewpoint. Section 3 describes the proposed analysis method. Section 4 presents the experiments and results. Finally, discussion and future work are drawn in Section 5.

2. Related Work

Dailey et al. [3] evaluated the effect of culture-specific facial expression understanding by analyzing the recognition capability of U.S. and Japanese participants. Their work is founded on a human study using a cross-cultural emotional expression dataset. In order to explore the interaction of the assumed universal expressions with cultural learning, they proposed to model the behavior of the human study by using a computational model based on Gabor filtering, PCA and perceptron artificial neural networks. Dailey's experiment helps to demonstrate how the interaction with other people in a cultural context defines the way of recognizing a culture-specific facial expression dialect. In summary, they found in-group advantages for recognizing facial expressions, since each racial group was better than the other at classifying facial expressions posed by members of the same culture.

In a more recent study, Jack et al. [4] claimed to refute the universal hypothesis of facial expressions by using generative grammars and visual perception for analyzing the mental representations of Western and Eastern cultural individuals. In this proposal, they model the facial expression representations per culture based on the 6 basic emotions and they found that each emotion is not expressed using a combination of facial movements common to both racial groups. Finally, Jack et al. concluded that the 6 basic emotions can clearly represent the Western facial expressions, but those are inadequate to accurately represent the conceptual space of emotions for East Asians, demonstrating a different cultural-specific, and not a universal representation of the basic emotions.

In summary, the mentioned cross-cultural studies have found differences on representing and categorizing facial expressions between cultures, concluding that facial expressions could be defined as culture-specific instead of universal. However, these findings are approached from a psychological viewpoint and did not consider the differences that can be found from automatic FER systems, including those which are known to appear in specific facial regions. Therefore an analysis which can cover those issues may serve to facial expressions understanding and will help to develop better automatic systems for facial expression and emotion recognition.

3. Proposed Method

The performance of any FER system depends on its feature extraction method, which is usually based on appearance or geometric features. Appearance features represent the skin texture of the face and its changes, meanwhile geometric features represent the shape of the face and facial parts. On this category, besides it is an old algorithm, PCA has repeatedly proved its efficiency as feature extraction method as well as for feature analysis [6]. Therefore, PCA is applied for extracting the appearance and geometric features, and it is the stand of our hybrid proposal.

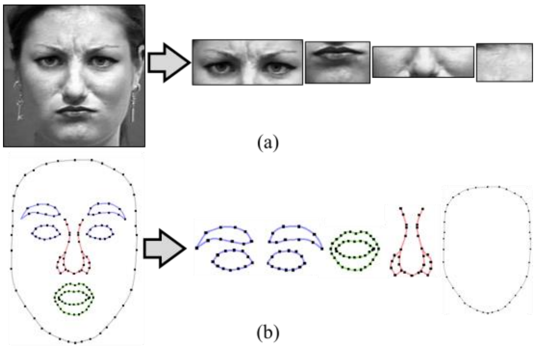

Figure 1. Facial segmentation, (a) appearance- and (b) geometric-based method.

3.1. Appearance-based method

For this method we use an algorithm proposed in [7]. Similar to the Eigenfaces algorithm which characterizes the face in an n-dimensional vector by applying PCA to the training set for obtaining the "Eigenspace" which is new space where projections made on it contains the most relevant texture information from the original face. This algorithm generates individual projections of four specific facial regions (eyes-eyebrows, nose, mouth and forehead). Therefore, independent Eigenspaces per each facial region are generated. The automatic segmentation of this method is based on the distance between irises and its relation with the rest of facial parts. Figure 1 (a) shows an example of the facial region segmentation.

3.2. Geometric-based method

A total of 163 feature points is used for defining the whole facial shape and the same process of individual PCA application for independent facial regions is applied. Therefore, the shape is segmented as shown in Figure 1(b), where 54 feature points define the shape of eyes-eyebrows, 42 for lips, 29 for nose and 38 for face outline. It is important to mention that, this paper intends to precisely analyze facial expression differences between two racial groups, thus the total number of feature points is large and those were manually obtained from each face image. Thereby, the accuracy of feature extraction is increased and the analysis capability enhanced. We have also developed an automatic extraction method of a large number of feature points, yet the results tend to be affected by shooting conditions and it is still not easy to guarantee the sufficient accuracy for the analysis.

3.3. Hybrid method

As mentioned before, the proposed hybrid method is based on PCA. Thus, firstly consider V_{geo}, a vector representing coordinates from the shape of a geometric-based facial region, and Es_{geo} which represents the Eigenspace of all individual V_{geo} from the same region. Hence, y_{geo} is the projection vector of the initial V_{geo}, but for our method, 100% of the variance has to be retained in the PCA process. Therefore, the size of y_{geo} is equal to that of V_{geo}. Simultaneously we apply PCA to appearance-based facial region, where V_{ape} and Es_{ape} represent the pixels vector and the Eigenspace, respectively. Next, y_{ape} which is the projection vector of V_{ape} has to be calculated with just the 90% of the variance, providing a vector significant smaller than V_{ape}. Subsequently, y_{geo} and y_{ape} have to be concatenated to conform H_{vec} which represents a hybrid vector of both different types of features. Finally, a new Eigenspace, Es_{reg}, is calculated using all of H_{vec} vectors from the current facial region. Thereby, the final feature vector is represented by Y_{vec} which is a projection from H_{vec} retaining the 99% variance of Es_{reg}. It is worth noting that, this is also the retained variance used for previously appearance- and geometric-based methods. Figure 2 shows the process for the calculation of a hybrid feature vector from a mouth region.

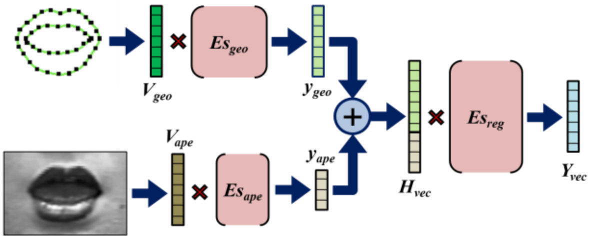

Figure 2. Process of hybrid feature vectors calculation.

4. Experiments and Results

Experiments presented in this paper were evaluated using the same datasets. The six basic expressions were classified by using multi-class SVMs with RBF kernels [8] and evaluated by leave one subject out cross-validation (only for automatic FER systems).

4.1. Datasets

Two datasets divided by Western (WSN) and East-Asian (ASN) people were conformed. The WSN dataset contains 240 facial images (40 per expression) selected from the Extended Cohn-Kanade database [9]. Meanwhile, the ASN dataset is comprised of 240 images (40 per expression) from, JAFFE database [10], JACFEE database [11] and TFEI database [12]. The face images were pre-processed to have the same inter-ocular distance and eye position, as well as cropped with 280x280 pixels.

4.2. Human study

A cross-cultural human study is presented as baseline result. This experiment was applied to 40 participants (20 Westerns and 20 East-Asians students) who were divided into two groups for answering a survey based on each dataset. We developed a computational program to collect a forced-choice facial expression classification from each participant using a whole dataset as stimuli. Each stimulus appeared in the central visual field and remained visible just for 3 seconds, followed by the 6-way forced choice decision question based on the prototypic basic expressions. Results of this study are presented in the following sections, depending on the in- or out-group experiment.

Table 1. Average recognition rate (%) of individual facial regions and their best combinations.

	Appearance		*Geometric*		*Hybrid*	
	WSN	*ASN*	*WSN*	*ASN*	*WSN*	*ASN*
Eye	68.3	65.8	62.1	61.7	74.2	68.8
Mou	84.6	73.3	90.4	80.4	95.8	81.7
Nos	66.3	46.3	75.4	55.4	81.7	61.3
FrOt	34.2	32.9	69.2	60.0	71.7	62.1
EMN	91.7	89.2	97.1	92.9	99.2	93.8
All	87.9	87.1	95.0	91.7	97.1	92.5

4.3. In-group analysis

In-group analysis represents the performance of FER when the people from the same race attempt to recognize facial expressions from their own racial group. In terms of FER systems, this happens when the training and testing sets correspond to the same dataset (WSN or ASN). Table 1 shows the results of the different facial regions using the 3 feature extraction methods divided by the dataset employed for the evaluation. *FrOt* refers to the forehead region or the outline shape as it may apply, and *EMN* represents the combination of eyes-eyebrows, mouth and nose regions. From Table 1 we can notice that the best results are provided by the proposed hybrid method. In addition, the WSN test reaches higher accuracy than the ASN for all feature extraction methods and facial regions.

Table 2 presents the performance per basic expression obtained by human study and the most relevant individual facial regions using the hybrid method. The results show that differences of FER between WSN and ASN are remarkable for eyes and mouth regions, especially fear on the eyes, and disgust on the mouth. In addition, the human study presents strong deficiency for recognizing fear.

Table 2. Recognition rate (%) per expression of in-group analysis from hybrid and human tests.

	Human Test		*Eye-Hybrid*		*Mou-Hybrid*	
	WSN	*ASN*	*WSN*	*ASN*	*WSN*	*ASN*
Ang	80.0	79.0	72.5	77.5	95.0	90.0
Dis	72.0	62.0	82.5	72.5	92.5	72.5
Fea	28.0	47.0	32.5	75.0	95.0	60.0
Hap	100	90.0	87.5	50.0	92.5	82.5
Sad	84.0	59.0	75.0	42.5	100	87.5
Sur	97.0	93.0	95.0	95.0	100	97.5
Avg	76.8	71.7	74.2	68.8	95.8	81.7

4.4. Out-group analysis

The performance of out-group FER test is based in how subjects can recognize facial expressions from a different racial group. For the proposed method, the testing set was the opposite dataset than that of training.

Table 3. Recognition rate (%) per expression of out-group analysis from hybrid and human tests.

	Human Test		*Eye-Hybrid*		*Mou-Hybrid*	
	WSN	*ASN*	*WSN*	*ASN*	*WSN*	*ASN*
Ang	49.0	74.0	60.0	52.5	90.0	77.5
Dis	58.0	44.0	42.5	57.5	52.5	62.5
Fea	28.0	49.0	30.0	50.0	65.0	82.5
Hap	97.0	92.0	77.5	70.0	60.0	85.0
Sad	70.0	51.0	35.0	30.0	75.0	80.0
Sur	97.0	93.0	82.5	92.5	97.5	87.5
Avg	66.7	67.2	54.6	58.8	73.3	79.2

Table 3 shows the results of out-group analysis of the human study, plus the eyes and the mouth regions from the hybrid method. Datasets cited in the table refer to the testing sets. Interesting results are obtained from this test. Westerns can categorize the faces of ASN dataset closer to the level reached by East-Asians. Comparing the results of Tables 2 and 3 we can see that the performance of proposed FER system presents similar characteristics to the human study. However, it is worth noting that in the out-group analysis, some expressions of ASN are better recognized than those of the in-group. That is the case of fear on the mouth region and happiness on the eyes. On the other hand, this phenomenon does not occur for the WSN test. This suggests that the 6 basic expressions fit better for WSN and overlap for some of ASN dataset.

Table 4. Average recognition rate (%) of the proposed hybrid method trained with both datasets (COM).

	Testing Set:		
	WSN	*ASN*	*COM*
Eye	73.8	62.9	68.3
Mou	89.6	85.0	87.3
EMN	97.1	92.9	95.0

Another way for the out-group analysis is when the system is trained with a combined dataset (COM) which includes both cultural datasets. Table 4 presents the results of this test using the hybrid method. Comparing the results of Tables 1 and 3 we can notice that the accuracy decreases when COM is used for training, suggesting that it is better to use a cultural specific training for improving the general performance of the FER system.

4.5. Visual analysis

Principal component scores of each projection vector represent the most valuable information from the face. Thus, for a visual analysis of these scores from geometric features, we used the Drawface tool [13] which shows the behavior of PC scores by drawing a caricatured face generated from a shape vector projected on a previously defined Eigenspace. Figures 3 and 4 show the caricatured faces (geometric-based) and the reconstructed images (appearance-based) obtained from averaged projection vectors per each basic expression. The first and second rows show the datasets of WSN and ASN respectively. Columns from left to right present the expressions of anger, disgust, fear, happiness, sadness and surprise.

From Figures 3 and 4 we can observe that most significant differences among WSN and ASN lie into the contrast of the expressions of disgust and fear. These findings could explain the performance of some results from Tables 2 and 3. In Table 2 for example, the low accuracy of fear on the WSN eyes region may be because of the similarity that this expression has in this region with sadness and anger. On the other hand, from the out-group analysis, the misrecognition of disgust on the WSN mouth region is completely understandable because disgusted ASN mouth looks entirely different than that of WSN.

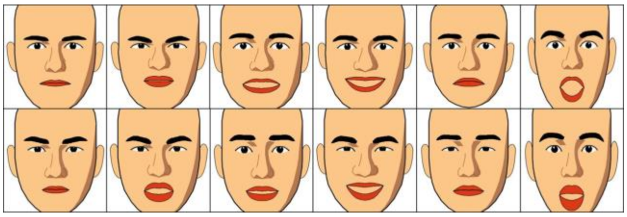

Figure 3. Caricaturized faces of averaged basic expressions (vectors) of WSN and ASN datasets.

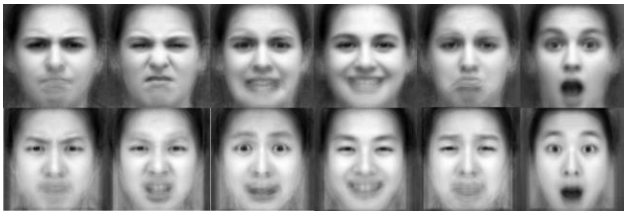

Figure 4. Image reconstructions of averaged basic expressions (vectors) of WSN and ASN datasets.

5. Discussion and Future Work

In this paper, we presented an analysis of culturally specific facial expression recognition focused on Western and East-Asian expressive faces. Despite of the limited database (40 faces per each basic expression), our study gives important discussion points. The way of showing facial expressions of Westerns and East-Asians is different, especially for disgust and fear on the regions of mouth and eyes-eyebrows. FER system recognizes better the prototypic expressions when WSN dataset is used for training (in-group analysis), and the highest results of the out-group analysis are obtained when the ASN dataset is used for test, these findings are comparable with the results of the human study. The proposed hybrid method is suitable for this analysis because it includes relevant information from shape and texture features of each facial region. Finally, even there exist newer feature extraction methods, PCA provides a deep expression analysis and a straight link possibility to visualize features by its principal components scores. In conclusion, the proposed method is just the beginning of a new way to analyze cultural differences of facial expressions approached from automatic FER systems.

As a future work, we plan to expand the limited size of the datasets by including databases from different countries (e.g. China and Korea). In addition, we plan to analyze the categorization capability of the 6 prototypic expressions among the racial groups by applying unsupervised classification methods to each dataset, thus it will enable the possibility to measure the cultural-specificity of the assumed basic expressions.

References

[1] C. Darwin: "The expression of the emotions in man and animals," Oxford University Press, 1872.

[2] P. Ekman: "An argument for basic emotions," *Cogn. Emotion*, vol. 6, pp. 169-200, 1992.

[3] M. N. Dailey, et al.: "Evidence and a computational explanation of cultural differences in facial expression recognition," *Emotion*, vol. 10, pp. 874-897, 2010.

[4] R. E. Jack, et al.: "Facial expressions of emotion are not culturally universal," *Proc. Natl. Acad. Sci.*, vol. 109, pp. 7241-7244, 2012.

[5] Y. Tian, et al.: "Facial Expression Recognition," in *Handbook of face recognition*, Springer, 2011, pp. 487-519.

[6] A. J. Calder, et al.: "A principal component analysis of facial expressions," *Vision Res.*, vol. 41, pp. 1179-1208, 2001.

[7] G. Benitez-Garcia, et al.: "Facial expression recognition based on facial region segmentation and modal value approach," *IEICE Trans. Inf. Syst.*, vol. 97, pp. 928-935, 2014.

[8] C. C. Chang and C. J. Lin: "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, paper number 27, 2011.

[9] T. Kanade, et al.: "Comprehensive database for facial expression analysis," *Proc. 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition* (*FG* 2000), pp. 46-53, 2000.

[10] M. Lyons, et al.: "Coding facial expressions with Gabor wavelets," *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition* (*FG'*98), pp. 200-205, 1998.

[11] M. Biehl, et al.: "Matsumoto and Ekman's JACFEE: Reliability data and cross-national differences," J. Nonverbal Behavior, vol. 21, pp. 3-21, 1997.

[12] L. F. Chen and Y. S. Yen: "Taiwanese facial expression image database," *Brain Mapping Laboratory, Institute of Brain Science*, Taiwan, 2007.

[13] M. Kaneko: "Computerized facial caricatures," *Journal of ITE*, vol. 62, pp. 1938-1943, 2008.