DOI: 10.1117/1.JEI.24.2.023015

Journal of Electronic Imaging 24(2), 023015 (Mar/Apr 2015)

Effects of cultural characteristics on building an emotion classifier through facial expression analysis

Flávio Altinier Maximiano da Silva and Helio Pedrini*
University of Campinas, Institute of Computing, Campinas, SP 13083-852, Brazil

Abstract. Facial expressions are an important demonstration of humanity's humors and emotions. Algorithms capable of recognizing facial expressions and associating them with emotions were developed and employed to compare the expressions that different cultural groups use to show their emotions. Static pictures of predominantly occidental and oriental subjects from public datasets were used to train machine learning algorithms, whereas local binary patterns, histogram of oriented gradients (HOGs), and Gabor filters were employed to describe the facial expressions for six different basic emotions. The most consistent combination, formed by the association of HOG filter and support vector machines, was then used to classify the other cultural group: there was a strong drop in accuracy, meaning that the subtle differences of facial expressions of each culture affected the classifier performance. Finally, a classifier was trained with images from both occidental and oriental subjects and its accuracy was higher on multicultural data, evidencing the need of a multicultural training set to build an efficient classifier. *© 2015 SPIE and IS&T* [DOI: 10.1117/1.JEI.24.2.023015]

Keywords: facial expressions; emotion recognition; multicultural training.

Paper 14501 received Aug. 20, 2014; accepted for publication Feb. 27, 2015; published online Mar. 19, 2015.

1 Introduction

The automated recognition of emotions[1–6] through computational analysis is a challenging task in the field of computer vision. Facial expressions[7–9] are fundamental in this sense since they are one of the main signs of humanity's emotions. The development of a precise and fast system, capable of identifying emotion through facial expression analysis, could be useful in many knowledge domains such as image retrieval, human-computer interfaces,[10,11] and action recognition, among others.[12] Machine learning algorithms are the most common approach to this matter; they have become popular due to their efficiency in reaching satisfactory results.[13–15]

This work investigates how the classifier system, trained with facial expression static images from one specific culture, reacts on classifying a test set of a different culture. Specifically, the classifier should cross-classify elements from predominantly American datasets with a predominantly Japanese one. The main contribution of our computational analysis is to demonstrate whether emotions are expressed by the same facial expressions in different cultures.

As many studies suggest,[16–18] there are "universal" facial expressions for specific emotions. In this sense, there is the standardization known as Facial Action Coding System, which is able to link facial expressions to emotions.[19,20] Ekman[18] points out that this evidence of expression universality is stronger for happiness, sadness, surprise, anger, disgust, and fear. For that reason, our experiment focuses on investigating only those emotions and the facial expressions associated with them. Such studies also conclude that although cultures may use the same facial expressions for specific emotions, what actually triggers them may vary from culture to culture. That is why our experiment uses no emotion triggering, but instead uses datasets of facial expression images already cataloged by people from their own culture.

Dailey et al.[21] performed a similar experiment. They concluded that the subtle differences in cultural manifestation of emotion are enough to confuse the algorithm; it is then necessary to build an efficient emotion detection system to train it to deal with these minor differences. Gabor filters[22] were used on every image to highlight the edges and textures of the face, and extracted feature vectors were used to train a neural network (NN) learning algorithm.[23]

In our experiments, we tested not only Gabor filters for image description, but also the histogram of oriented gradients (HOGs) filter[24] and the local binary patterns (LBPs) filter.[5,25] For classifiers training, we used support vector machines (SVMs),[26] neural networks (NNs),[23] and *k*-nearest neighbors (k-NNs).[27]

This paper is organized as follows. Section 2 presents some concepts and works related to the topic under investigation. Section 3 describes our proposed methodology. Section 4 presents and discusses some of the obtained results with the proposed method. Finally, Sec. 5 concludes our work and includes some future work suggestions for improving the proposed method.

2 Background

Human emotion recognition has received increasing attention in several knowledge domains such as action recognition, human-computer interactions, behavior prediction, affective computing, and health care, among several others. Emotion is a subjective experience or a physiological reaction of human beings[7] that can be demonstrated in the form of facial expressions, voice intonation, hand gestures, and body language.

*Address all correspondence to: Helio Pedrini, E-mail: helio@ic.unicamp.br

1017-9909/2015/$25.00 © 2015 SPIE and IS&T

A nonverbal and universal communication attribute that describes the emotions in all human beings is facial expressions. Two main theories have been formulated to define the concept of emotion in the psychological field. Discrete theory[28] has been developed by psychologists to describe emotions based on the hypothesis that there exist universal basic emotions. Ekman,[16–18] for instance, conducted several studies to support the idea that the emotion perception in different cultures is present in six basic facial expressions (anger, disgust, fear, happiness, sadness, and surprise). On the other hand, dimensional theory[29,30] describes the emotions in terms of small sets of dimensions, which include control, power, evaluation, and activation, among others.

Several works[2–4,6–8,10,11,14,15,31] have been developed by the scientific community to automatically recognize facial expressions. Some methods are based on the recognition of emotions, whereas others are based on the recognition of facial muscle actions or facial action units. The facial action units describe signals that can be translated into emotion categories through high-level mapping.

Most of the available approaches are based on a number of two-dimensional spatiotemporal facial features.[5,6,32–34] Such features are commonly categorized into appearance and geometric features. Appearance features attempt to represent facial texture characteristics such as protuberances, wrinkles, or furrows. Geometric features attempt to capture the shape of facial components such as mouth, nose, chin, and eyes. Methods to automatically recognize expressions based on three-dimensional face models have also been developed.[35–38]

More recent works have investigated the problem of expression analysis of videos.[39–41] To deal with such a dynamic process, the methods must be capable of effectively considering temporal alignment and semantic representation.[40]

In order to promote and improve the development of automatic expression recognition approaches, several challenges[42,43] and datasets[44–51] have been created which aim to establish a common platform for creating and validating expression recognition methods in both controlled and real-world conditions.

A summary of some relevant results obtained with state-of-the-art methods on four public datasets is presented in the tables. It is worth mentioning that many approaches adopted different protocols for the same data. Furthermore, some methods applied specific preprocessing stages to the data such as alignment or cropping of the images and intensity adjustments for reducing the influence of lighting conditions, among others.

Table 1 reports the results for the Extended Cohn–Kanade (CK+) dataset.[44,45] Wang et al.[52] performed a 15-fold cross-subject validation, Littlewort et al.[53] took a subset of the dataset for evaluation, Chew et al.[54] adopted a leave-one-subject-out cross-validation, and Jain et al.[55] employed a fourfold cross-validation, whereas Sanin et al.[56] adopted a fivefold cross-validation. The results shown for the methods developed by Scovanner et al.,[57] Wang et al.,[58] Zhao and Pietikainen,[59] and Klaser and Marszalek[60] were obtained with the same data and protocols used by Liu et al.[40]

Table 2 shows the results for the Japanese Female Facial Expression (JAFFE) dataset.[46] Lyons et al.[62] performed a 10-fold cross-validation and Liang et al.[63] divided the set into two equal parts for training and testing, whereas Shinohara and Otsu,[64] Zheng et al.,[65] Xue and Youwei,[66] Horikawa,[67] Kyperountas et al.,[68] Feng et al.,[69] He et al.,[70] Gu et al.,[71] Xue and Gertner,[72] and Wang et al.[73] adopted a leave-one-subject-out cross-validation.

Table 3 presents the results for the MUG Facial Expression Database (MUG) dataset.[48] Rahulamathavan et al.[75] and

Table 1 Accuracy rates (in percentage) for CK+ dataset.

Method	Strategy	Accuracy (%)
Scovanner et al.[57]	Three-dimensional scale-invariant feature transform (3-D SIFT)	81.35
Zhao and Pietikainen[59]	Local binary patterns on three orthogonal planes (LBP-TOP)	88.99
Klaser and Marszalek[60]	Histograms of oriented 3-D spatiotemporal gradients (HOG 3-D)	91.44
Lucey et al.[45]	Active appearance models (AAM)	83.30
Littlewort et al.[53]	Computer expression recognition toolbox (CERT)	87.21
Ptucha et al.[61]	Manifold-based sparse representation (MSR)	91.40
Jain et al.[55]	Temporal modeling of shapes (TMS)	91.89
Chew et al.[54]	Modified correlation filters (MCF)	89.40
Sanin et al.[56]	Spatiotemporal covariance (Cov3D)	92.30
Wang et al.[58]	Histogram of spatiotemporal orientation energy (HOE)	82.26
Wang et al.[52]	Interval temporal Bayesian network (ITBN)	86.30
Liu et al.[40]	Spatiotemporal manifold (STM)	91.13

Table 2 Accuracy rates (in percentage) for JAFFE dataset.

Method	Strategy	Accuracy (%)
Lyons et al.[62]	Gabor filters and linear discriminant analysis (LDA)	75.00
Shinohara and Otsu[64]	Higher-order local autocorrelation (HLAC) and Fisher weight maps	69.40
Feng et al.[69]	Local binary patterns (LBP)	77.00
Liang et al.[63]	Supervised locally linear embedding (SLLE)	79.54
He et al.[70]	Enhanced local binary patterns (LBP)	79.21
Zheng et al.[65]	Kernel canonical correlation analysis (KCCA)	77.05
Xue and Youwei[66]	Difference of statistical features (DSF)	62.78
Horikawa[67]	Kernel canonical correlation analysis (KCCA) and Kansei information	67.00
Wang et al.[73]	Locality-preserved maximum information projection (LPMIP)	83.18
Kyperountas et al.[68]	Salient feature vectors (SFVs)	85.92
Thai et al.[74]	Canny edge detector and artificial neural networks	85.70
Gu et al.[71]	Radial encoded Gabor features	89.67
Xue and Gertner[72]	Gaussian pyramid decomposition and Gabor wavelet filter	92.90

Aina et al.[76] adopted a leave-one-out cross-validation. Table 4 reports the results for the BOSPHORUS 3D Face Database (BOSPHORUS) dataset.[47] Savran and Sankur,[77] Zhao et al.,[78] and Savran et al.[79] adopted a 10-fold cross-validation.

For additional details on concepts and works related to expression recognition, we refer the reader to some surveys.[12,36,80–83]

3 Methodology

The main goal of our methodology is to investigate, using a variety of filters and machine learning algorithms, whether a multiclass classifier is capable of correctly classifying emotions on images of cross-cultural facial expressions. As the literature suggests,[16–18] six main emotions are very similar in all studied cultures. However, as seen in Dailey et al.,[21] in a first analysis, the classifier did not perform so well.

Four different public datasets were used in our experiments. For predominantly occidental images, the evaluated datasets were the CK+,[44,45] the MUG,[48] and BOSPHORUS.[31,47,84] The chosen oriental dataset was the JAFFE.[46]

The emotions considered relevant for this study were happiness, sadness, surprise, anger, disgust, and fear. For each dataset, only the images labelled with such emotions were considered. For the CK+ dataset, 309 images were analyzed; for MUG, 376; for BOSPHORUS, 453; and for JAFFE, 182.

To guarantee the uniformity in the images studied, all pictures were cropped and resized to 96×96 pixels. CK+, MUG, and BOSPHORUS datasets provide the facial landmarks annotations for each image; these landmarks were used in the cropping process so that only the facial characteristics were taken into account. Since the JAFFE dataset does not provide such information, all its images were cropped manually.

As each dataset also employs a different lighting scheme for the images, a histogram equalization technique was used on every image—as literature suggests, this process should make cross-database classification more efficient. Specifically,

Table 3 Accuracy rates (in percentage) for MUG dataset.

Method	Strategy	Accuracy (%)
Rahulamathavan et al.[75]	Local fisher discriminant analysis (LFDA)	95.24
Aina et al.[76]	Eigenfaces and sparse representation-based classification (SRC)	91.27

Table 4 Accuracy rates (in percentage) for BOSPHORUS dataset.

Method	Strategy	Accuracy (%)
Savran and Sankur[77]	Action unit (AU) detection	91.40
Zhao et al.[78]	Extended statistical facial feature models (SFAM)	94.20
Savran et al.[79]	Action unit (AU) detection with fusion of 2-D and 3-D data	97.10

Fig. 1 Examples of three samples used in our experiments. From left to right: sample from the JAFFE dataset that represents anger; sample from CK+ that represents disgust; and sample from BOSPHORUS that represents happiness.

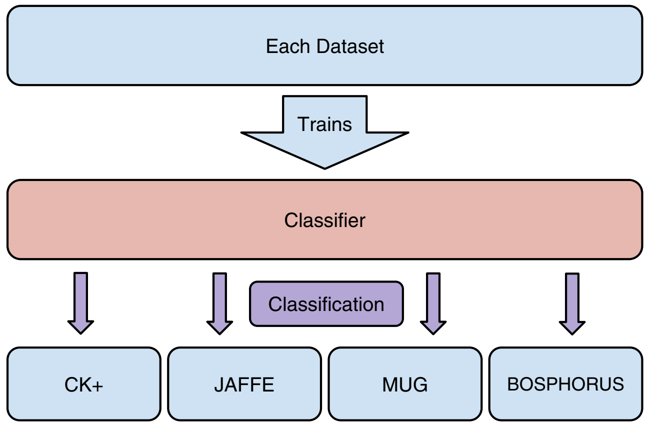

Fig. 2 Diagram with the main steps of the proposed methodology. Classifiers were trained using each one of the four datasets and were then used to classify them.

the filter used in this stage was the exact histogram specification.[85] Figure 1 shows three examples of the datasets, before and after cropping, rescaling, and histogram equalization procedures.

Three different filters were used on the images to detect which would result in the best accuracy for these datasets. The evaluated filters were the HOG,[24] Gabor filters,[22] and LBPs.[5,25]

For HOG, images were divided into a 9×9 grid, then the oriented gradients within each square formed the feature vectors, which would later be fed to the classifier. Thus each image was represented by a 81-dimensional feature vector.

The Gabor filters used 5 scales in 8 orientations, with 39 rows and columns. Rows and columns were downsampled by a factor of 4. This produces a feature vector of approximately 22,000 features for each image; for a large number of images, the classifiers would take too much time to be trained and tested. For this reason, and to guarantee a fair comparison with HOG (which produces a 81-dimensional feature vector for each image), the 22,000-dimensional feature vectors were also reduced to 81 dimensions. The technique used for this dimensionality reduction was principal component analysis (PCA).[86]

The tested LBP filter used a radius of 1 pixel for every pixel in each image. The resulting feature vector was composed of more than 9000 features. As performed for the Gabor filtered data, these results were reduced to 81 features through PCA. Therefore, every image was represented by three 81-dimensional feature vectors—one for HOG, one for Gabor, and the other for the LBP results.

Furthermore, three different classification algorithms were evaluated: SVM, NNs, and k-NN. The datasets were partitioned as follows: for each facial expression, approximately 60% of its elements were used to train the classifier; 20% of the other elements were used as a cross-validation group (used to calibrate the parameters of the classifier); and the remaining 20% of the data were used as the test group. The only exception is k-NN, where the cross-validation sets were also used for training.

Each dataset was used to train three emotion classifiers (one for each learning algorithm). These models were then used to classify not only the dataset which trained them, but also the others. In this sense, we could determine whether classifiers trained in one dataset generalized more accurately for datasets from the same cultural group rather than on those from different cultural groups. When testing on a dataset different from the one that trained the classifier, all images from this dataset were used as a test group since the classifier never had contact with any of them.

For SVM, library for support vector machines[87] was employed in our implementation; in particular, the radial basis function kernel was chosen. With this kernel, two parameters must be calibrated by using the cross-validation group. Twelve different values were chosen and tested for each, which resulted in 144 different combinations. Only the one with the best results over the cross-validation group was used to classify the test group.

For NN, 1 hidden layer with 10 hidden neurons composed the net—the MATLAB built-in functions were used for this purpose. Also, MATLAB functions were used to train and test the k-NN algorithms; tests with $k = 1, 3, 5, 7$ were performed; however, as $k = 1$ presented the best results, only those are reported in the experiment section.

Figure 2 shows a diagram illustrating the steps taken in this experiment. To summarize, each combination of description filter and learning algorithm was used 16 times—one for each combination of datasets.

After that, a classifier was trained with both the CK+ and MUG datasets, representing a more general occidental classifier (since the BOSPHORUS dataset classifiers produced very little accuracy, we decided not to include this dataset in the occidental classifier). It was then used to classify the JAFFE dataset (oriental) and we analyzed the model accuracy on each emotion. A classifier trained with the JAFFE dataset was also applied on this occidental dataset of CK+ and MUG.

Finally, a multicultural classifier was built from joining together CK+, MUG, and JAFFE datasets. The accuracy for this model was analyzed when classifying multicultural data for every emotion.

4 Experimental Results

As expected, for most combinations of descriptor and learning algorithms, the in-group classification was satisfactory; however, when classifying different datasets, the accuracy usually drops considerably. Before proceeding to more detailed tests, we decided to choose the combination of filter and classifier that gives the best overall accuracy.

In this sense, an analysis of each combination must be performed. To choose the combination for more detailed tests, we proceed as follows: evaluation of the classifiers trained on one dataset and tested on the same dataset; the combination of filter and classifier with best and most consistent accuracy will be chosen as the best combination.

Table 5 summarizes the results of such a combination. It is possible to see that the results for each combination vary; however, some patterns can be observed. For example, every combination of filter and classifier yields poor accuracy on the BOSPHORUS dataset. BOSPHORUS is a dataset with many complications, such as beards and moustaches, which may confuse the algorithm. For this reason, we have decided not to use this dataset to build a more "general" occidental database—only CK+ and MUG were used for this. The results for this combination of datasets are given as "MUG + CK+" in Table 5.

Another interesting fact is the superiority of the HOG and Gabor filters over LBP in almost every combination. LBP has high accuracy only in combinations where HOG and Gabor filters also have high accuracies. We conclude that, for the given objectives of this experiment, LBP is not an adequate filter.

Table 5 Accuracy is given, in percentage, for each combination of filter, learning algorithm, and dataset. The best accuracy for the combination is given in bold.

		Classifier		
Dataset	Filter	SVM	NN	k-NN
CK+	HOG	84.10	76.09	66.67
	Gabor	87.30	**93.48**	46.03
	LBP	79.36	82.61	49.21
JAFFE	HOG	80.50	66.67	**91.67**
	Gabor	75.00	88.89	77.78
	LBP	66.66	40.74	66.67
MUG	HOG	81.00	75.00	89.19
	Gabor	81.08	**89.29**	89.19
	LBP	82.43	85.71	87.84
BOSPHORUS	HOG	63.33	67.16	27.78
	Gabor	62.22	**73.13**	14.44
	LBP	54.44	61.19	31.11
MUG + CK+	HOG	**84.70**	75.49	79.56
	Gabor	76.64	66.67	65.69
	LBP	69.34	56.86	67.15

It is also possible to notice that the combination of the Gabor filter and NNs provided the best results on most datasets; however, it performed quite poorly for the combination of the CK+ and MUG datasets. As the focus of this experiment is to test cultural databases, in particular the "MUG + CK+" dataset, it would not be a wise choice for this experiment.

For this reason, we decided to conduct the experiments by using the HOG filter and SVM as the learning algorithm. This combination provided the most consistent results—almost always around 80%, with the only exception being the BOSPHORUS dataset.

As expected and shown in Table 5, the in-group classification (with HOG filter and SVM model) was satisfactory, resembling results from the literature. All cross-classification accuracies are shown for the HOG filter and SVM classifier in Table 6.

It is clear that when the model is used to classify the same dataset as the one it was trained with, the accuracy is higher than when it is used on other datasets. As a rule, the accuracy drops severely in those cases. The only exception is the BOSPHORUS dataset, where the accuracy was low in all conducted tests. For this reason, the BOSPHORUS dataset was not considered in further analysis. It is also possible to see that, although the models trained with CK+ and MUG datasets have low accuracy when classifying the other occidental datasets, the accuracy is still higher than when classifying JAFFE, the oriental one. This could indicate that, beyond the physical differences between the datasets, such as lightning conditions, there must be something else that reduces the accuracy when they are applied on JAFFE—this could be the cultural factor.

A deeper analysis for each emotion on these cross-classification tests is as follows. The results for each dataset are shown in the following confusion matrices, where the column headings specify the correct classification, and the row headings the actual classification (where Hap stands for happiness, Sad for sadness, Sur for surprise, Ang for anger, Dis for disgust, and Fea for Fear). The cell values represent how many times the classifier guessed wrong and, in the main diagonal, the percentage of correct classifications. The $\sum$ column shows the number of times an emotion was incorrectly chosen, and the $\sum$ row the number of times each emotion was not recognized. The $(\sum, \sum)$ cell gives the total accuracy (in percentage). Finally, the "Total" row shows the

Table 6 All 16 combinations of cross-classification between the four datasets. The datasets shown on the left are the ones used to train the classifier. The cells give the accuracy, in percentage, of that classifier when applied to the dataset on the column header.

	Test set			
Training set	CK+	JAFFE	MUG	BOSPHORUS
CK+	84.1	42.3	47.8	43.0
JAFFE	48.2	80.5	32.9	30.0
MUG	45.6	32.4	81.0	53.8
BOSPHORUS	57.6	36.2	56.6	63.3

Table 7 Confusion matrix for the JAFFE dataset. Classification of 35 samples.

	Hap	Sad	Sur	Ang	Dis	Fea	$\sum$
Hap	83.3%	0	0	0	0	0	0
Sad	1	83.3%	0	0	1	0	2
Sur	0	0	100.0%	0	0	0	0
Ang	0	0	0	83.3%	0	0	0
Dis	0	0	0	1	80%	0	1
Fea	0	1	0	0	0	100.0%	1
$\sum$	1	1	0	1	1	0	88.6%
Total	6	6	6	6	5	6	35

Table 9 Confusion matrix for the MUG dataset. Classification of 74 samples.

	Hap	Sad	Sur	Ang	Dis	Fea	$\sum$
Hap	88.2%	0	0	0	2	0	2
Sad	0	100.0%	0	1	0	1	2
Sur	0	0	84.6%	1	0	1	2
Ang	1	0	1	72.7%	1	2	5
Dis	1	0	0	0	78.6%	0	1
Fea	0	0	1	1	0	55.6%	2
$\sum$	2	0	2	3	3	4	81.1%
Total	17	10	13	11	14	9	74

total number of samples in each category, whereas the rightmost cell shows the total number of elements studied. For example, the cell (Ang, Dis) in Table 8 shows that one sample that should be classified as disgust was actually classified as anger.

Tables 7, 8, and 9 show the in-group classification for the JAFFE, CK+, and MUG datasets, respectively. For the JAFFE dataset, the training group had 112 samples, whereas the cross-validation group had 36. For the CK+ dataset, 183 samples were used for training and 63 for validation. For MUG, the classifier was trained with 228 samples and validated with 74.

It is possible to observe for the predominantly oriental dataset (JAFFE) that the results are fairly homogeneous when comparing the emotions. For the occidental datasets, CK+ and MUG, on the other hand, the classifier performed quite poorly on some emotions and very well on others. We assumed this is due to the number of training samples: on the CK+ dataset, the classifier was trained with 43 samples of happiness and was very effective in classifying it. For fear, however, there were only 15 training examples and it could recognize only two of three test samples. Similarly, on MUG, the classifier was trained with 53 samples of happiness, while only 29 samples of fear were present. Sadness seems to be the most distinctive emotion, since on both datasets there are very few test samples and the classifier had perfect accuracy on them.

Before proceeding to cross-cultural tests, a deeper analysis of the association of MUG and CK+ datasets was conducted. A classifier was trained with both of the training groups from MUG and CK+ and validated with both of their cross-validation groups. Table 10 shows the confusion matrix for the classification over the association of their test sets.

It is possible to see that the pattern of the occidental datasets is maintained: high accuracy on all emotions except for anger and fear. The overall accuracy also had a small increase when compared with results when the classifiers were trained on individual datasets and tested on them.

It would, therefore, seem then fair to use the classifier trained on eastern faces—which had a homogeneous amount

Table 8 Confusion matrix for the CK+ dataset. Classification of 63 samples.

	Hap	Sad	Sur	Ang	Dis	Fea	$\sum$
Hap	85.7%	0	0	1	0	1	2
Sad	0	83.3%	0	0	0	0	0
Sur	0	0	88.2%	0	0	1	1
Ang	0	0	0	77.8%	1	0	1
Dis	0	1	0	0	91.7%	0	1
Fea	2	0	2	1	0	60.0%	5
$\sum$	2	1	2	2	1	2	84.1%
Total	14	6	17	9	12	5	63

Table 10 Confusion matrix for the "MUG + CK+" dataset.

	Hap	Sad	Sur	Ang	Dis	Fea	$\sum$
Hap	90.3%	0	0	1	1	1	3
Sad	0	93.8%	0	3	0	0	3
Sur	0	0	90.0%	1	0	2	3
Ang	0	0	1	70.0%	2	2	5
Dis	1	1	0	0	88.5%	0	2
Fea	2	0	2	1	0	64.3%	5
$\sum$	3	1	3	6	3	5	84.7%
Total	31	16	30	20	26	14	137

Table 11 Confusion matrix for the classifier trained with the oriental dataset (JAFFE) used to label the occidental one (MUG + CK+).

	Hap	Sad	Sur	Ang	Dis	Fea	$\sum$
Hap	44.2%	1	20	9	38	4	72
Sad	8	67.1%	3	35	12	8	66
Sur	1	4	58.4%	5	11	18	39
Ang	12	13	25	23.5%	21	14	85
Dis	42	2	9	18	15.4%	6	77
Fea	24	5	5	11	28	30.6%	73
$\sum$	87	25	62	78	110	50	39.9%
Total	156	76	149	102	130	72	685

of training samples for each emotion—on the occidental datasets. This time, we could use the entire CK+ and MUG datasets as the test set, as the classifier had never had any contact with them. The results of this experiment are shown in Table 11.

As can be seen in Table 11, the results are not as good as expected. The classifier had a strong drop on every emotion when compared with the previous tests. Sadness was the only emotion with some reasonable classification results—it is possible to observe from both Tables 7 and 10 that the results of the sadness classification were also very good. At this point, we can assume that, through this classification approach, sadness is a fairly easy emotion to classify and might be considered universal. The worst case is on happiness and disgust: it classified 42 samples of happiness as disgust and 38 samples of disgust as happiness. It seems to confuse those two emotions quite a lot—giving evidence that there might be some intercultural correlation between them. In general, the accuracy was very poor: only 39.9%.

Table 12 Confusion matrix for the classifier trained with the occidental dataset (MUG + CK+) used to label the oriental one (JAFFE).

	Hap	Sad	Sur	Ang	Dis	Fea	$\sum$
Hap	58.1%	4	1	3	3	3	14
Sad	0	16.1%	1	3	2	4	10
Sur	1	2	83.3%	6	0	3	12
Ang	1	14	0	36.7%	9	11	35
Dis	7	6	0	4	37.9%	5	22
Fea	4	0	3	3	4	16.1%	14
$\sum$	13	26	5	19	18	26	41.2%
Total	31	31	30	30	29	31	182

Table 13 Confusion matrix for the multicultural classifier trained with MUG, CK+, and JAFFE datasets.

	Hap	Sad	Sur	Ang	Dis	Fea	$\sum$
Hap	83.8%	1	0	1	2	1	5
Sad	0	68.2%	0	2	0	4	6
Sur	0	0	83.3%	1	0	1	2
Ang	0	3	2	80.8%	2	0	7
Dis	2	3	0	1	84.4%	0	6
Fea	4	0	4	0	1	70.0%	9
$\sum$	6	7	6	5	5	6	79.8%
Total	37	22	36	26	32	20	173

We then experimented on the contrary: using the classifier trained with predominantly occidental subject images to label the oriental dataset. The results can be seen in Table 12, where the overall accuracy is slightly better: 41.2%. Here, the surprise classification showed the best performance—one should conclude from this fact that the surprise facial expression must be universal. The classifier confusion between happiness and disgust is also still present, although less significant. It seems to confuse anger and sadness as well—most of the sadness samples were incorrectly classified as anger. When we turn our attention back to Table 11, we see this confusion is also present: most anger samples were classified as sadness. Therefore, it is clear that the classifier is not suitable for intercultural classification.

Ekman,[18] however, strongly concludes that the six basic emotions are universal. Our hypothesis is that the subtle intercultural differences are enough to confuse the classifier; we then tried to train a classifier with both cultural bases. "MUG + CK+" and JAFFE datasets were all combined into one multicultural dataset and used to train a classifier. If the problem was caused by the subtle differences, then since the classifier had contact with those datasets, it should be able to distinguish them more accurately. The result of this multicultural classification can be seen in Table 13.

A general improvement can be seen for every emotion—aside from sadness—and the classification accuracy is much more homogeneous when comparing each emotion. It is possible to see that the apparent confusion between happiness and disgust is much more subtle. The same goes for the confusion between sadness and anger. In fact, the confusion between emotions is well distributed, and we attribute this variance not only to cultural differences, but also to physical discrepancies between the datasets themselves such as lighting conditions, for example.

5 Conclusions and Future Work

Evidence from our experiments suggest that the six basic emotions are universal with a few minor differences. Classifiers trained with a multicultural dataset performed well on multicultural test groups—there are still confusions between some facial expressions, though they could be influenced not only by cultural differences, but also by other aspects of the datasets such as lighting. Classifiers trained with single-culture data performed poorly on the other culture data.

We believe that the minor differences between facial expressions in different cultures are enough to confuse the classifier. In this sense, we reinforce that a vast dataset of multicultural samples is needed to build truly efficient emotion detection systems through facial expression analysis.

Directions for future work include the use of larger datasets for similar experiments. We also intend to study how classifiers react to partial occlusion of the facial expression samples, as well as techniques for overcoming such obstacles.

Acknowledgments

The authors are grateful to FAPESP—São Paulo Research Foundation (Grants 2011/22749-8 and 2014/04020-9) and CNPq (Grant 307113/2012-4) for their financial support.

References

1. V. Gay et al., "CaptureMyEmotion: helping autistic children understand their emotions using facial expression recognition and mobile technologies," *Stud. Health Technol. Inf.* **189**, 71–76 (2013).
2. A. Majumder, L. Behera, and V. K. Subramanian, "Emotion recognition from geometric facial features using self-organizing map," *Pattern Recognit.* **47**(3), 1282–1293 (2014).
3. S. Ryoo and J.-K. Chang, "Emotion affective color transfer using feature based facial expression recognition," *Adv. Sci. Technol. Lett.* **39**, 131–135 (2013).
4. J. Rojas, A. Ramirez, and O. Chae, "Facial expression recognition based on local sign directional pattern," in *19th IEEE Int. Conf. Image Processing*, pp. 2613–2616, IEEE, Orlando, FL (2012).
5. C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image Vision Comput.* **27**(6), 803–816 (2009).
6. T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," in *27th Conf. Image and Vision Computing*, pp. 464–468, ACM, Dunedin, New Zealand (2012).
7. S. Zhang, X. Zhao, and B. Lei, "Facial expression recognition using sparse representation," *WSEAS Trans. Syst. Control* **11**, 440–441 (2012).
8. S. Rifai et al., "Disentangling factors of variation for facial expression recognition," *Lect. Notes Comput. Sci.* **7577**, 808–822 (2012).
9. M. Mortillaro, M. Mehu, and K. R. Scherer, *The Evolutionary Origin of Multimodal Synchronization and Emotional Expression. Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, Oxford University Press, New York, NY (2013).
10. Z. Zeng et al., "A survey of affect recognition methods: audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(1), 39–58 (2009).
11. G. Giorgana and P. Ploeger, "Facial expression recognition for domestic service robots," *Lect. Notes Comput. Sci.* **7416**, 353–364 (2012).
12. M. F. Valstar et al., "The first facial expression recognition and analysis challenge," in *IEEE Int. Conf. Automatic Face & Gesture Recognition*, pp. 921–926, IEEE, Santa Barbara, CA (2011).
13. M. S. Bartlett et al., "Machine learning methods for fully automatic recognition of facial expressions and facial actions," in *IEEE Int. Conf. Systems, Man and Cybernetics*, Vol. 1, pp. 592–597, IEEE, The Hague, Netherlands (2004).
14. L. Ma and K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," *IEEE Trans. Syst., Man, Cybern., B* **34**(3), 1588–1595 (2004).
15. O. Pheobe, H. DongJun, and R. Rimiru, "Machine learning performance on face expression recognition using filtered backprojection in DCT-PCA domain," *Int. J. Comput. Sci. Issues* **10**(1), 145 (2013).
16. P. Ekman, "Universal and cultural differences in facial expressions of emotion," in *Nebraska Symp. Motivation*, University of Nebraska Press, Lincoln, NE (1971).
17. P. Ekman, "Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique," *Psychol. Bull.* **115**(2), 268–287 (1994).
18. P. Ekman, "Facial expressions," in *Handbook of Cognition and Emotion*, T. Dalgleish and M. J. Power, Eds., Vol. 16, pp. 301–320, John Wiley & Sons, Ltd., Chichester, UK (1999).
19. J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the facial action coding system," in *The Handbook of Emotion Elicitation and Assessment*, James A. Coan and John J.B. Allen, Eds., pp. 203–221, Oxford University Press, New York (2007).
20. P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System Investigator's Guide, A Human Face*, Research Nexus, Salt Lake City, Utah (2002).
21. M. N. Dailey et al., "Evidence and a computational explanation of cultural differences in facial expression recognition," *Emotion* **10**(6), 874–893 (2010).
22. I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biol. Cybern.* **61**, 103–113 (1989).
23. M. N. Dailey et al., "EMPATH: a neural network that categorizes facial expressions," *J. Cognit. Neurosci.* **14**(8), 1158–1173 (2002).
24. O. Ludwig, Jr. et al., "Trainable classifier-fusion schemes: an application to pedestrian detection," in *12th Int. IEEE Conf. Intelligent Transportation Systems*, pp. 432–437, IEEE, St. Louis, Missouri (2009).
25. T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," in *European Conf. Computer Vision*, Vol. 35, pp. 469–481, Springer, Prague, Czech Republic (2004).
26. S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, Burlington, MA (2009).
27. G. Shakhnarovich, T. Darrell, and P. Indyk, *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice (Neural Information Processing)*, The MIT Press, Cambridge, MA (2006).
28. A. Ortony and T. Turner, "What's basic about basic emotions?" *Psychol. Rev.* **97**(3), 315–331 (1990).
29. J. Russell, L. Ward, and G. Pratt, "Affective quality attributed to environments: a factor analytic study," *Environ. Behav.* **13**, 259–288 (1981).
30. K. Scherer, "What are emotions? And how can they be measured?," *Soc. Sci. Inf.* **44**(4), 695–729 (2005).
31. S. B. A. Savran and T. Bilge, "Comparative evaluation of 3D versus 2D modality for automatic detection of facial action units," *Pattern Recognit.* **45**, 767–782 (2012).
32. T. R. Almaev and M. F. Valstar, "Local Gabor binary patterns from three orthogonal planes for automatic facial expression recognition," in *Humaine Association Conf. Affective Computing and Intelligent Interaction*, pp. 356–361, IEEE, Geneva, Switzerland (2013).
33. S. M. Lajevardi and Z. M. Hussain, "Automatic facial expression recognition: feature extraction and selection," *Signal, Image Video Process.* **6**(1), 159–169 (2012).
34. S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Comput. Vision Image Understanding* **115**(4), 541–558 (2011).
35. Y. Chang et al., "Automatic 3D facial expression analysis in videos," in *IEEE Int. Workshop Analysis and Modeling of Faces and Gestures*, Vol. 3723, pp. 293–307, IEEE, Beijing, China (2005).
36. T. Fang et al., "3D facial expression recognition: a perspective on promises and challenges," in *IEEE Int. Conf. Automatic Face & Gesture Recognition and Workshops*, pp. 603–610, IEEE, Santa Barbara, CA (2011).
37. G. Sandbach et al., "A dynamic approach to the recognition of 3D facial expressions and their temporal models," in *IEEE Int. Conf. Automatic Face & Gesture Recognition and Workshops*, pp. 406–413, IEEE, Santa Barbara, CA (2011).
38. J. Wang et al., "3D facial expression recognition based on primitive surface feature distribution," in *IEEE Int. Conf. Computer Vision and Pattern Recognition*, Vol. 2, pp. 1399–1406, IEEE, New York, NY (2006).
39. X. Huang et al., "Expression recognition in videos using a weighted component-based feature descriptor," in *Image Analysis*, pp. 569–578, Anders Heyden and Fredrik Kahl, Eds., Springer, Berlin, Heidelberg (2011).
40. M. Liu et al., "Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1749–1756, IEEE, Columbus, OH (2014).
41. T. Robin, M. Bierlaire, and J. Cruz, "Dynamic facial expression recognition with a discrete choice model," *J. Choice Modell.* **4**(2), 95–148 (2011).
42. Facial Expressions in the Wild, "Emotion recognition in the wild challenge and workshop," 2014, http://cs.anu.edu.au/few/emotiw2014.html (14 August 2014).
43. European Network of Excellence in Social Signal Processing, "Facial expression recognition and analysis challenge," 2015, http://sspnet.eu/fera2015/ (5 January 2015).
44. T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Fourth IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 46–53, IEEE, Grenoble, France (2000).
45. P. Lucey et al., "The extended Cohn-Kanade Dataset (CK+): a complete facial expression dataset for action unit and emotion-specified expression," in *Third IEEE Workshop on CVPR for Human Communicative Behavior Analysis*, San Francisco, California (2010).
46. Database of Digital Images, http://www.kasrl.org/jaffe.html (14 August 2014).
47. A. Savran et al., "Bosphorus database for 3D face analysis," in *Biometrics and Identity Management*, Ben Schouten et al., Eds., pp. 47–56, Springer, Berlin Heidelberg (2008).
48. N. Aifanti, C. Papachristou, and A. Delopoulos, "The MUG facial expression database," in *11th Int. Workshop on Image Analysis for*

Multimedia Interactive Services, pp. 1–4, IEEE, Desenzano del Garda, Italy (2010).
49. A. Dhall et al., "Collecting large, richly annotated facial-expression databases from movies," *IEEE Multimedia* **19**, 34–41 (2012).
50. M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: an addition to the MMI facial expression database," in *Int. Conf. Language Resources and Evaluation, Workshop on EMOTION*, pp. 65–70, European Language Resources Association, Valletta, Malta (2010).
51. G. Zhao et al., "Facial expression recognition from near-infrared videos," *Image Vision Comput.* **29**, 607–619 (2011).
52. Z. Wang, S. Wang, and Q. Ji, "Capturing complex spatio-temporal relations among facial muscles for facial expression recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 3422–3429, IEEE, Portland, OR (2013).
53. G. Littlewort et al., "The computer expression recognition toolbox (CERT)," in *IEEE Int. Conf. Automatic Face Gesture Recognition and Workshops*, pp. 298–305, IEEE, Santa Barbara, CA (2011).
54. S. W. Chew et al., "Improved facial expression recognition via uni-hyperplane classification," in *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2554–2561, IEEE, Providence, RI (2012).
55. S. Jain, C. Hu, and J. K. Aggarwal, "Facial expression recognition with temporal modeling of shapes," in *IEEE Int. Conf. Computer Vision Workshops*, pp. 1642–1649, IEEE, Barcelona, Spain (2011).
56. A. Sanin et al., "Spatio-temporal covariance descriptors for action and gesture recognition," in *IEEE Workshop on Applications of Computer Vision*, pp. 103–110, IEEE, Clearwater Beach, FL (2013).
57. P. Scovanner, S. Ali, and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," in *15th ACM Int. Conf. Multimedia*, pp. 357–360, ACM, Augsburg, Germany (2007).
58. L. Wang, Y. Qiao, and X. Tang, "Motionlets: mid-level 3D parts for human motion recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2674–2681, IEEE, Portland, OR (2013).
59. G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(6), 915–928 (2007).
60. A. Klaser, M. Marszałek, and C. Schmid, "A spatio-temporal descriptor based on 3D-gradients," in *19th British Machine Vision Conf.*, Vol. 275, p. 1, British Machine Vision Association, Leeds, UK (2008).
61. R. Ptucha, G. Tsagkatakis, and A. Savakis, "Manifold based sparse representation for robust expression recognition without neutral subtraction," in *IEEE Int. Conf. Computer Vision Workshops*, pp. 2136–2143, IEEE, Barcelona, Spain (2011).
62. M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(12), 1357–1362 (1999).
63. D. Liang et al., "A facial expression recognition system based on supervised locally linear embedding," *Pattern Recognit. Lett.* **26**(15), 2374–2389 (2005).
64. Y. Shinohara and N. Otsu, "Facial expression recognition using Fisher weight maps," in *Sixth IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 499–504, IEEE, Seoul, Republic of Korea (2004).
65. W. Zheng et al., "Facial expression recognition using kernel canonical correlation analysis (KCCA)," *IEEE Trans. Neural Networks* **17**(1), 233–238 (2006).
66. G. Xue and Z. Youwei, "Facial expression recognition based on the difference of statistical features," in *8th Int. Conf. Signal Processing*, Vol. 3, IEEE, Beijing, China (2006).
67. Y. Horikawa, "Facial expression recognition using KCCA with combining correlation kernels and Kansei information," in *Int. Conf. Computational Science and its Applications*, pp. 489–498, Springer-Verlag, Kuala Lumpur, Malaysia (2007).
68. M. Kyperountas, A. Tefas, and I. Pitas, "Salient feature and reliable classifier selection for facial expression classification," *Pattern Recognit.* **43**(3), 972–986 (2010).
69. X. Feng, A. Hadid, and M. Pietikäinen, "A coarse-to-fine classification scheme for facial expression recognition," in *Image Analysis and Recognition*, Aurélio Campilho and Mohamed Kamel, Eds., pp. 668–675, Springer, Berlin Heidelberg (2004).
70. L. He et al., "An enhanced LBP feature based on facial expression recognition," in *27th Annual Int. Conf. Engineering in Medicine and Biology Society*, pp. 3300–3303, IEEE, Shanghai, China (2005).
71. W. Gu et al., "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," *Pattern Recognit.* **45**(1), 80–91 (2012).
72. H. Xue and I. Gertner, "Automatic recognition of emotions from facial expressions," *Proc. SPIE* **9090**, 90900O (2014).
73. H. Wang et al., "Locality-preserved maximum information projection," *IEEE Trans. Neural Networks* **19**(4), 571–585 (2008).
74. L. H. Thai, N. D. T. Nguyen, and T. S. Hai, "A facial expression classification system integrating canny, principal component analysis and artificial neural network," *Int. J. Mach. Learn. Comput.* **1**, 388–393 (2011).
75. Y. Rahulamathavan et al., "Facial expression recognition in the encrypted domain based on local Fisher discriminant analysis," *IEEE Trans. Affective Comput.* **4**(1), 83–92 (2013).
76. S. Aina et al., "A new spontaneous expression database and a study of classification-based expression analysis methods," in *22nd European Signal Processing Conference*, IEEE, Lisbon, Portugal (2014).
77. A. Savran and B. Sankur, "Automatic detection of facial actions from 3D data," in *IEEE 12th Int. Conf. Computer Vision Workshops*, pp. 1993–2000, IEEE, Kyoto, Japan (2009).
78. X. Zhao et al., "AU recognition on 3D faces based on an extended statistical facial feature model," in *Fourth IEEE Int. Conf. Biometrics: Theory Applications and Systems*, pp. 1–6, IEEE, Washington, DC (2010).
79. A. Savran, B. Sankur, and M. T. Bilge, "Facial action unit detection: 3D versus 2D modality," in *IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops*, pp. 71–78, IEEE, San Francisco, CA (2010).
80. A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: a survey," *IEEE Trans. Affective Comput.* **4**(1), 15–33 (2013).
81. S. Mahto and Y. Yadav, "A survey on various facial expression recognition techniques," *Int. J. Adv. Res. Electr., Electron. Instrum. Eng.* **3**, 13028–13031 (2014).
82. V. J. Mistry and M. M. Goyani, "A literature survey on facial expression recognition using global features," *Int. J. Adv. Eng. Technol.* **2**, 653–657 (2013).
83. G. Sandbach et al., "Static and dynamic 3D facial expression recognition: a comprehensive survey," *Image Vision Comput.* **30**(10), 683–697 (2012).
84. A. Savran, B. Sankur, and M. T. Bilge, "Regression-based intensity estimation of facial action units," *Image Vision Comput.* **30**(10), 774–784 (2012).
85. D. Coltuc, P. Bolon, and J.-M. Chassery, "Exact histogram specification," *IEEE Trans. Image Process.* **15**(5), 1143–1152 (2006).
86. I. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York (2002).
87. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**, 27:1–27:27 (2011).

Flávio Altinier Maximiano da Silva is currently a computer engineering student at the University of Campinas, Brazil. His research interests include machine learning, computer vision, pattern recognition, image processing, and computer graphics.

Helio Pedrini is currently a professor in the Institute of Computing at the University of Campinas, Brazil. He received his PhD degree in electrical and computer engineering from Rensselaer Polytechnic Institute, Troy, New York, United States. He received his MSc degree in electrical engineering and his BSc degree in computer science from the University of Campinas, Brazil. His research interests include image processing, computer vision, pattern recognition, machine learning, computer graphics, and scientific visualization.