Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/090550

Cross-Cultural Emotion Classification based on Incremental Learning and LBP-Features

M. Sultan Zia[1] *and M. Arfan Jaffar*[1,2,*]

[1] National University of Computer and Emerging Sciences, Islamabad, Pakistan
[2] College of Computer and Information Sciences, Al-Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

Received: 15 Feb. 2015, Revised: 17 May 2015, Accepted: 18 May 2015
Published online: 1 Sep. 2015

Abstract: A number of studies have shown that facial expression representations are cultural dependent and not universal. Most facial expression recognition (FER) systems use one or two datasets for training and same for testing and show good results. While their performance mortify radically when datasets from different cultures were presented. To keep high accuracy for a long time and for all cultures, a FER system should learn incrementally. We proposed a FER system that can offer incremental learning capability. Local Binary Pattern (LBP) Features are used for Region of Interest (ROI) extraction and classification. We used static images of facial expressions from different cultures for training and testing. The experiments on five different datasets using the incremental learning classification demonstrate promising results.

Keywords: Incremental Learning, Classification, Facial Expression, LBP, Template Matching

1 Introduction

Facial expression recognition (FER) has a wide range of applications like human-computer interaction, virtual reality, video conferencing and synthetic face animation etc. However, automatic facial expression recognition is still a challenging problem due to illumination, head pose, aging, glasses, cultures and races. Even, human cannot 100% recognize facial expressions due to many reasons. First, different people interpret expressions differently for example, fear expression images, which interpreted as fear in America and most of the other countries are mostly interpreted as surprise in Japan. Second, people from different cultures think about facial expressions in different ways, for example, in some cultures, people focus on mouth but in others focus on eyes etc. Third, in different cultures expressions are expressed in different ways i.e. not all the facial expressions are innate and universal. Figure 1 shows the cultural variations in representation of two facial expressions. Row 1 shows some variations of disgust expression while row 2 shows some variations of surprise expression. Similarly cultural variations for some other expressions exist as well. Many studies and surveys like [1,2,3,4,5,6] have shown that many facial expressions are cultural dependent and not innate or universal. The problem of cultural variations in facial expressions is a well-studied problem in psychology literature but it is never discussed and/or considered (as per our knowledge) in information sciences. Therefore, this factor is never kept in mind while developing facial expression recognition systems.

Conventional expression recognition systems show brilliant performance when they are tested over the datasets on which they were trained. But on the other hand, when they are plow into the practical environment, their performance mortify drastically. Such performance degradation is due to the reason that the facial expression dataset used while training of the classifier, was either insufficient or inappropriate for future uses. Even if a large dataset of expression images is available for training, it seems impossible that it can deal with all the variations that could ever happen in future. Therefore, it is hardly to be expected that a system trained on static dataset can show high performance in practical situations. Another reason of the failure of such systems is the fact that most of the benchmark datasets composed of posed (not spontaneous) expression images/videos. In psychology it is a well-known fact posed and spontaneous expressions are different in their temporal dynamics, characteristics and timings.

* Corresponding author e-mail: arfan.jaffar@ccis.imamu.edu.sa

In this paper, we proposed a solution to such like problems by using the incremental learning concept. We did this by embedding an incremental learning ability into the facial expression classification. Learn++ is such an algorithm that have ability to learn incrementally [9].

Fig. 1: Cutural variations in facial expressions are shown. All images in row 1 represents disgust expressions and all images in row 2 represents surprise expression

The other promising benefit of using incrementally learning classification systems is the reduction of variance and increase in confidence of the decision. Moreover, due to many random variations in a given classifier model (different training data, different initialization, etc.), the decision obtained by any given classifier may vary substantially from one training trial to another even if the model structure is kept constant. Then, combining the outputs of several such classifiers can reduce the risk of an unfortunate selection of a poorly performing classifier. To the best of our knowledge the problem of cross cultural emotion expressions is not considered yet. Only one paper [7] found where across the datasets experiments were conducted. Our results are much better than [7].

2 Proposed Method

We present Local Binary Pattern Histogram (LBPH) Features [8] to an incrementally learning algorithm which uses Template Matching as base classifier. The block diagram of our proposed approach is given in figure 2.

In order to recognize the expression from a static image of an individual, our region of interest (ROI) is the face of that individual. Local Binary Pattern (LBP) features have performed very well in various applications. The Extended LBP operator is denoted by $LBP^{U2}_{P,R}$ where P is the number of neighboring pixels, R is radius of neighbor circle and $U2$ for uniform patterns. We used $LBP^{U2}_{P,R}$ for expression classification and face detection. Texture classification method proposed in [8] on the bases of LBP features is used to extract the region of interest. Let $D_d = \{I_{d,1}, I_{d,2}, I_{d,3}, \ldots I_{d,md}\}$ be the set of face images in each dataset d, where md stands for number of images in dataset d. In this paper we used five different

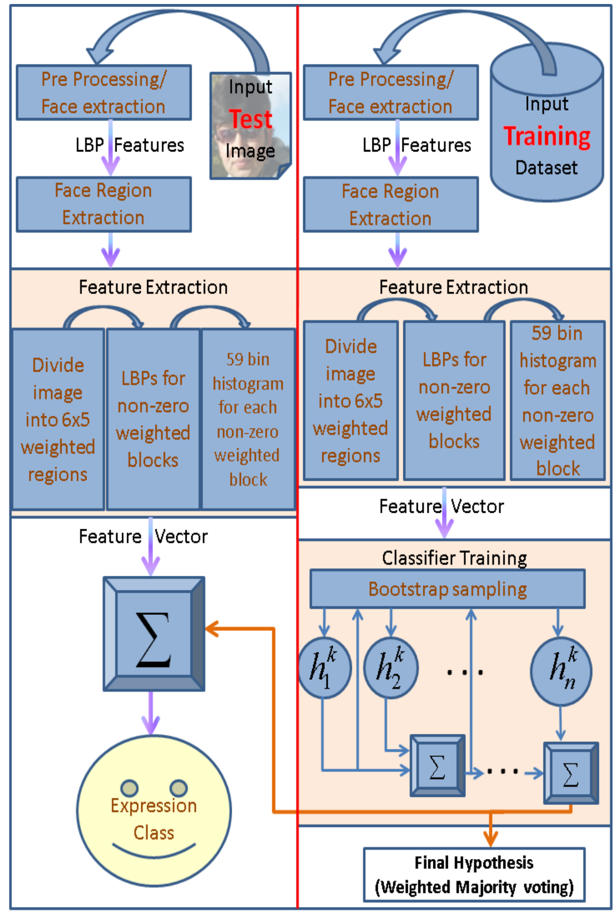

Fig. 2: Block diagram of the proposed system: right side represents the training process and left side represent the testing process

datasets i.e. $d = 1, 2, 3, \ldots, 5$. For supervised incrementally learning algorithms, the class label of $I_{d,i}$ is assumed to be $l_i \in \{1, 2, \ldots, c\}$ where c is the number of classes, 7 in our case.

For each image, in each dataset the image is divided into 5x6 weighted regions as shown in figure 3. The resulting image is denoted by $j^{r,c}_{d,i} = div(I_{d,i})$, and the corresponding LBPH features for each region is computed as follows:

$$h^{r,c}_{d,i} = LBPH(J^{r,c}_{d,i}) \tag{1}$$

where $r = 1, 2, \ldots, 6$ and $c = 1, 2, \ldots, 5$. The size of each $h^{r,c}_{d,i}$ is 1x59. There are 30 (5x6) such histogram features for each image $I_{d,i}$. Then a concatenated histogram of size 1x1770 is computed.

$$H_{d,i} = NORM\left(\left[h^{r,c}_{d,i}\right]^{r=6,c=5}_{r=1,c=1}\right) \tag{2}$$

This feature vector is presented to an incremental learning algorithm Learn++ [9]. We used Template Matching as base/weak classifier. By combining the outputs of several so-called weak classifiers by, for example, averaging the output decisions, we can reduce the risk of an unfortunate selection of a poorly performing classifier [9,10]. However, a more interesting and arguably more challenging problem is the introduction of new classes or different number of classes being represented in each new dataset. For expression recognition systems, it is a practical problem. By making strategic modifications to the bootstrap resampling distribution, a similar approach can still be used to learn incrementally under these scenarios [11]. Learn++ is such an algorithm shown to learn incrementally from new data, even when such data introduce new classes [11]. To train the classifier on some dataset d, the dataset is divided into training and testing subsets. To construct a reference template histogram, an initial uniform distribution (without any prior knowledge) is assigned to the training subset. A bootstrap training sample S is obtained. A reference histogram template is computed for each class C by averaging all $h_{d,i}$ features in S belonging to class C.

$$T_i^c = \frac{1}{N_{s,c}} \sum_{j=1}^{N_{s,c}} h_{d,j}^c \tag{3}$$

where $N_{s,c}$ is the number of histograms in S belonging to class C. These templates can be expected to be near the class they belonging to and far from the remaining provided that the chi square dissimilarity measure could successfully discriminate samples belonging to different classes. Then we optimize these templates by using a Fischer's discriminate ratio for multiclasses.

$$R = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} \frac{(\mu_i - \mu_j)^2}{\sigma_i{}^2 + \sigma_j{}^2} \tag{4}$$

Where μ_i and σ_i are mean and standard deviation of samples' χ^2 distances from template T for i^{th} expression class. A method proposed by M.S.Zia et al. [12] is used to maximize the above ratio. Then an iterative method is applied to obtained optimal templates for each class. Then for each $(i+1)^{th}$ template the distribution is updated on the bases of previous i templates. A higher probability is assigned to misclassified expression images before selecting the $(i+1)^{th}$ bootstrap sample. Hence more and more focus on critical expression images. For detail referred to [9,11].

Then a nearest-neighbor classifier is used to match the input image with the closest template. We selected Chi square statistic χ^2 as the dissimilarity measure [13] for histograms.

$$\chi^2(S,M) = \sum_{j=1}^{B} \frac{(S_i - M_i)^2}{S_i + M_i} \tag{5}$$

where S and M are two LBP histograms and B is the number of bins (1770 in our case). It is observed that

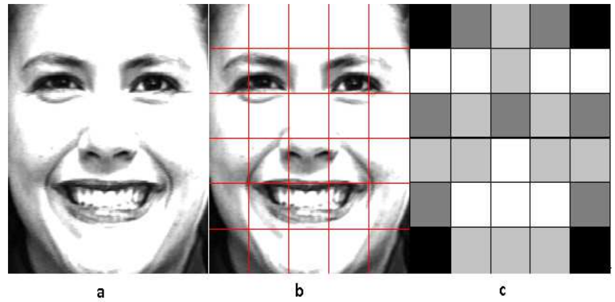

Fig. 3: (a) A face image (cohn-kanade), (b) face image divided into 5x6 sub-region, (c) The weights set for weighted dissimilarity measure. Black squares indicate weight 0, dark gray 1, light gray 2 and white 3.

facial features contributing to facial expressions mainly lie in some regions, such as eye area and mouth area; these regions contain more useful information for facial classification. Therefore, a weight can be set for each face region based on the importance of the information it contains (figure 3). The weighted χ^2 statistic is

$$\chi_w^2(S,M) = \sum_{1,2} w_j \frac{(S_{i,j} - M_{i,j})^2}{S_{i,j} + M_{i,j}} \tag{6}$$

where w_j is the weight for region j. The ensemble of many such base classifiers is combined through weighted majority voting [9] that represents the overall ensemble decision.

3 Experimentation

Five different datasets were used for the experimentation purposes. In each training session, only one of these datasets was used. For each training session, 20 weak hypotheses (TEMPLATE MATCHING) were generated by Learn++. Each hypothesis (histogram template) of the i^{th} training session was generated using a training subset and a testing subset (used to compute hypothesis error). Histogram Template matching is used as base classifier. An additional validation set, TEST, of 200 images was used for validation purposes. During incremental learning process the Learn++ tracks those images of the new dataset that have already been learned by the current ensemble, and instead focuses on other images that carry the novel information. Finally, note that all classifiers are retained to prevent loss of information, assuming that all previous data still carry relevant information. If previous information is no longer relevant, then a forgetting mechanism can be introduced to remove irrelevant base classifiers [10]. We train our incremental learning algorithm on the four datasets one by one and the fifth dataset namely TEST was used for evaluation purposes only. We used five different databases to evaluate the performance of our proposed approach to facial expression classification. Each of these datasets contains hundreds of images. The five datasets used in this study are JAFFE [14], Cohn Kanade [15], MUG [16], FEEDTUM [17] and our own dataset of colored images of Indian subcontinent people, movies and videos. Few sample face images from these datasets are shown in figure 4. The number of images corresponding to each of the 7 categories of expression (neutral, happiness, sadness, surprise, anger, disgust and fear) is roughly the same. We used 80% to 90% images of each dataset for training and remaining for testing. We generate a bootstrap sample of 98 to 175 images to train each of the (10 to 20) weak classifiers for each dataset.

Fig. 4: : Images from five datasets Row wise from top to bottom MUG, Cohn Kanade, FEED, JAFFE and Our Own dataset (TEST) respectively

To show and prove the soundness of incremental learning we perform a systematic experimentation. One by one datasets were presented to the Learn++ for training the classifiers and at each stage the performance of classification is measured over all the five datasets.First of all we train the incremental classifier by using JAFFE dataset only and call it $H_1(\mathbf{X})$. Then we evaluated the performance of classifier $H_1(\mathbf{X})$ on JAFFE along with the remaining four datasets. The results of this experimentation are shown in table 1. Now the previously trained classifier on $H_1(\mathbf{X})$ is provided with Cohn Kanade dataset for successive training. Note that in this training session only Cohn-Kanade dataset was presented to Learn++ for training, nor JAFFE neither the other three datasets. This classifier is called $H_2(\mathbf{X})$ and its performance is shown in table 2. Similarly we trained the classifier incrementally by using MUG and FEED datasets (one at a time) and the corresponding classifiers will be called $H_3(\mathbf{X})$ and $H_4(\mathbf{X})$ respectively. The results of the performance evaluation of $H_3(\mathbf{X})$ on all five datasets were shown in table 3, while the results of $H_4(\mathbf{X})$ on all five datasets were shown in table 4. Note that the overall performance of the classifier increased with successive training sessions, which shows that the algorithm has the ability to learn the variations in expressions present in the different datasets. Whether the variations were due to the cultural variations or due to different representation patterns of facial muscles.

Note that we did not train any classifier/hypothesis on the fifth dataset namely TEST. The performance on this dataset can be considered as the performance in real world.

Table 1: Accuracies in Percentage for different Expressions (Rows) and datasets (Columns) after training the Learn++ on JAFFE ($H_1(\mathbf{X})$)

Expr \ DS	JAF	C.K.	MUG	FEED	TEST
Ne	96.9	58.2	55.2	57.8	51.1
Ha	97.4	56.1	51.4	54.2	46.6
Sa	96.5	51.6	53.1	52.4	49.9
An	97.1	53.7	49.7	51.5	33.8
Su	97.2	54.3	52.8	48.3	50.3
Di	95.2	48.4	48.9	50.5	34.6
Fe	94.7	54.3	51.7	52.1	44.2
Avg	**96.4**	**53.8**	**51.8**	**52.4**	**44.3**

Table 2: Accuracies in Percentage for different Expressions (Rows) and datasets (Columns) after training the Learn++ on Cohn Kanade ($H_2(\mathbf{X})$)

Expr \ DS	JAF	C.K.	MUG	FEED	TEST
Ne	96.3	92.1	69.3	68.7	57.9
Ha	95.5	98.2	71.1	70.8	64.1
Sa	96.1	97.7	68.7	66.1	61.5
An	97.2	95.9	72.6	69.7	62.6
Su	97.5	94.8	73.4	71.6	64.5
Di	95.8	95.1	65.8	67.2	60.8
Fe	93.5	94.4	66.3	64.7	62.7
Avg	**95.9**	**95.4**	**69.6**	**68.4**	**62.0**

Caifeng et al. [13] proposed a method of facial expression recognition based on LBP features and they

Table 3: Accuracies in Percentage for different Expressions (Rows) and datasets (Columns) after training the Learn++ on MUG ($H_3(\mathbf{X})$)

Expr \ DS	JAF	C.K.	MUG	FEED	TEST
Ne	97.5	99.1	98.2	86.9	87.9
Ha	98.2	95.7	97.9	88.2	82.8
Sa	96.9	97.9	97.4	87.6	79.6
An	96.4	92.5	96.5	85.9	82.7
Su	100	93.7	98.6	96.8	80.6
Di	95.9	96.9	96.8	85.7	86.2
Fe	96.4	97.2	96.5	86.5	70.7
Avg	**97.3**	**96.1**	**97.4**	**88.2**	**81.5**

Table 4: Accuracies in Percentage for different Expressions (Rows) and datasets (Columns) after training the Learn++ on FEED ($H_4(\mathbf{X})$)

Expr \ DS	JAF	C.K.	MUG	FEED	TEST
Ne	100	97.2	96.7	96.3	90.1
Ha	98.5	98.1	97.9	98.7	89.4
Sa	96.8	97.1	97.5	97.7	85.8
An	97.9	98.2	98.8	98.6	83.3
Su	98.7	97.3	98.0	97.2	82.3
Di	95.4	96.7	96.9	95.9	87.5
Fe	96.1	96.0	95.5	95.8	85.7
Avg	**97.6**	**97.2**	**97.3**	**97.1**	**86.3**

used two classifiers namely, template matching and support vector machines. They also used Adaboost to find the boosted LBP features. In table 5, we compared the results of our proposed method with [13] on Cohn Kanade dataset [15].

Table 5: Comparison between the proposed method and Caifeng et al [13]

Methods (Feature + Classifier)	Recognition Results
LBP + Template Matching (TM)	79.1%
LBP + SVM(RBF)	87.6%
Boosted LBP based LDA	77.6%
Boosted LBP + SVM(RBF)	91.4%
Proposed (LBP + Learn++(TM))	**97.2%**

4 Conclusion

In conclusion we have proposed a method for automatic facial expression classification using incremental learning and local binary patterns histogram features. The proposed system has ability to learn incrementally and can accommodate future data. Due to incremental learning ability it can accommodate itself in different cultures. The experimentation performed is showing promising results.

Acknowledgement

The authors would like to thank Higher Education Commission (HEC), Govt. of Pakistan and NU-FAST, Islamabad for providing funds and required resources to complete this work.

References

[1] Jeanne L. Tsai, Yulia Chentsova-dutten, Variation among European Americans in emotional facial expression, journal of cross-cultural psychology, **34(6)**, 650-657 (2003).
[2] Elfenbein, H.A., Ambady, N., On the universality and cultural specificity of emotion recognition: A meta-analysis, Psychological Bulletin, **128**, 203235, (2002).
[3] Marsh, A., Elfenbein H.A., & Ambady N., Nonverbal accents: Cultural differences in facial expressions of emotion, Psychological Science, **14**, 373376, (2003).
[4] Matsumoto D., Cultural influences on the perception of emotion, Journal of Cross-Cultural Psychology, **20**, 92105, (1989).
[5] Matsumoto D., Methodological requirements to test a possible ingroup advantage in judging emotions across cultures: Comments on Elfenbein and Ambady and evidence, Psychological Bulletin, **128**, 236242, (2002).
[6] Rachael E. Jack, Roberto Caldara, and Philippe G. Schyns, Internal Representations Reveal Cultural Diversity in Expectations of Facial Expressions of Emotion, Journal of Experimental Psychology, **141(1)**, 1925, (2012).
[7] H. Yan, Marcelo H. Ang Jr, and A. N. Poo, Cross-Dataset Facial Expression Recognition, IEEE Int. Conf. on Robotics and Automation, May 9-13, (2011).
[8] T. Ojala and M. Pietikainen. Multiresolution, T., Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, **24**, 971987, (2002).
[9] R. Polikar, L. Udpa, S.S. Udpa, and V. Honavar, Learn++: An incremental learning algorithm for supervised neural networks, IEEE Trans. Systems, Man. Cybern. C, **31(4)**, 497508, (2001).
[10] Jun-Bao L., Shu-Chuan C., Jeng-Shyang P., Lakhmi C. J., Multiple Viewpoints Based Overview for Face Recognition, Journal of Information Hiding and Multimedia Signal Processing, **3**, 352-369, (2012).

[11] R. Polikar, Bootstrap-Inspired Techniques in Computational Intelligence, IEEE SIGNAL PROCESSING MAGAZINE, 59-72, (2007).
[12] M. S. Zia, and M. a. Jaffar, An adaptive training based on classification system for patterns in facial expressions using SURF descriptor templates, Multimedia Tools and Applications, 1-19, (2013).
[13] Caifeng S., Shaogang G., Peter W. M., Facial expression recognition based on Local Binary Patterns: A comprehensive study, Image and Vision Computing, **27(6)**, 803816, (2009).
[14] Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi & Jiro Gyoba, Coding Facial Expressions with Gabor Wavelets, Proceedings,Third IEEE ICAFGR, 200-205, (1998).
[15] Takeo Kanade Yingli Tian Jeffrey F. Cohn, Comprehensive Database for Facial Expression Analysis, Fourth IEEE ICAFGR, (2000).
[16] N. Aifanti, C. Papachristou and A. Delopoulos, The MUG Facial Expression Database, 11th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), 12-14, (2010).
[17] F. Wallhoff, FEEDTUM: Facial Expressions and Emotion Database, Technical University of Munich, (2005).

M. Sultan Zia received his B.Sc degree from Islamia University Bahawalpur, Pakistan in 2001. He later received Master in Computer Science (MCS) degree in 2005 from International Islamic University Islamabad, Pakistan. Then he received his M.S. degree in computer science in 2008 from National University of Computer and Emerging Sciences, NU-FAST, Islamabad, Pakistan. Currently he is pursuing his Ph.D. research at the Department of Computer Science, NU-FAST. His research interests include image processing, pattern recognition and computational intelligence.

M. Arfan Jaffar received his B.Sc degree from Bahauddin Zakariya University Multan, Pakistan in 2000 and got distinction (Gold Medal). He later received M.Sc. degree in computer science in 2003 from Quaid-e-Azam University Islamabad, Pakistan. Then he earned his PhD. degree in computer science in July 2009 from National University of Computer and Emerging Sciences, NU-FAST, Islamabad, Pakistan. Currently he is working as a Research Professor at Al-Imam Muhammad ibn Saud Islamic University, Riyadh, Saudi Arabia. His research interests include image processing, medical Imaging and computational intelligence.