FACIAL EXPRESSION RECOGNITION USING ADVANCED LOCAL BINARY PATTERNS, TSALLIS ENTROPIES AND GLOBAL APPEARANCE FEATURES

Shu Liao, Wei Fan, Albert C. S. Chung and Dit-Yan Yeung

Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Hong Kong.
{liaoshu,fwkevin,achung,dyyeung}@cse.ust.hk

ABSTRACT

This paper proposes a novel facial expression recognition approach based on two sets of features extracted from the face images: texture features and global appearance features. The first set is obtained by using the extended local binary patterns in both intensity and gradient maps and computing the Tsallis entropy of the Gabor filtered responses. The second set of features is obtained by performing null-space based linear discriminant analysis on the training face images. The proposed method is evaluated by extensive experiments on the JAFFE database, and compared with two widely used facial expression recognition approaches. Experimental results show that the proposed approach maintains high recognition rate in a wide range of resolution levels and outperforms the other alternative methods.

Index Terms— Face recognition, Machine vision,

1. INTRODUCTION

Within the last couple of years, automatic facial expression recognition (AFER) has gained an increasing interest in building natural human-computer interaction (HCI) systems. This task poses a significant challenge to the pattern recognition and computer vision research communities for the non-rigid deformations of different human expression styles as well as their large inter-personal variations. In this paper, we propose a novel approach to facial expression recognition from static images using elegantly extracted features which cover both textual and appearance characteristics of different expression patterns.

Two main strategies have been recently exploited by different AFER approaches [1]. Frame-based approaches use only the current frame with or without a reference image (in most cases corresponding to the neutral expression) to recognize the expression of the frame. On the other hand, sequence-based approaches use the temporal information to recognize various expressions for the short video clips. However, they are limited to the fact that the timing of facial actions of different expressions may be unstable when compared with their appearance configurations in each individual frame.

Frame-based AFER methods do not use temporal information for the recognition task but treat each input image as an independent sample of some fixed but unknown distributions. The key to their success lies on the extraction of expressive and discriminating image features which could be either obtained by certain predefined models or learned from large volume of training samples. Turk and Pentland [2] use Principle Component Analysis (PCA) to construct eigenfaces

and represent face images as projection coefficients along these basis directions. Gabor-wavelet features [3] and lower frequency 2-D DCT coefficients [4] are also explored in the literature. Apart from these widely used features, Local Binary Pattern (LBP) [6] as a novel low-cost image descriptor for texture classification has also been introduced to the field of facial expression analysis.

Although the conventional LBP approach [7] efficiently encodes the texture features of micro-pattern information in the face image, it only considers the uniform patterns. Our empirical findings show that the dominant patterns (i.e. the patterns with the largest proportion in the image) of some kinds of facial expressions are not uniform. In this case, the conventional LBP approach may not be able to reliably reflect the major structural information in the face images. Moreover, since the conventional LBP approach only extracts the high-frequency texture characteristics (i.e., pixel-wise relation) of the face images, the low- and mid-frequency texture features are therefore not effectively captured. Apart from the textural level information, the face images also contain appearance level features such as global facial motions, which could be well captured by linear subspace modeling (e.g., PCA and LDA).

Based on the above observations, we explore various feature extraction techniques for effectively characterizing different facial expression patterns. The main contributions of this paper are: (1) The conventional LBP approach is extended to the advanced LBP (ALBP) [8] in both intensity and gradient maps to ensure reliable acquisition of the major pattern information; (2) The Tsallis entropy [9] is used to extract the low- and mid-frequency texture features of the face image; (3) Discriminating global appearance features are extracted by null-space based LDA (NLDA) [10]. Then these three kinds of features are combined to represent the characteristics of the face image. The proposed method is evaluated by performing facial expression recognition using the JAFFE database [11] compared with two widely used methods. To observe its performance under low-resolution conditions, we also perform recognition experiments by reducing the resolution level of the original input images.

2. ADVANCED LOCAL BINARY PATTERNS

2.1. Advanced LBP in Intensity and Gradient Maps

In the conventional LBP approach [7], the image pixels are first labelled as a binary class by thresholding the difference between the center pixel and its neighbors using the step function $u(x)$ (i.e. $u(x) = 1$ when x≥0 and $u(x) = 0$ otherwise). The concatenation

of the neighboring labels is then used as a unique descriptor for each pattern. Figure 1 gives a simple example. The patterns are uniform if the transitions between "0" and "1" are less than or equal to two. For example, 01100000 and 11011111 are uniform patterns. The histogram of the uniform patterns in the whole image is used as the feature vector [7]. It has been proven to be effective for both face recognition and facial expression recognition applications [12] [6].
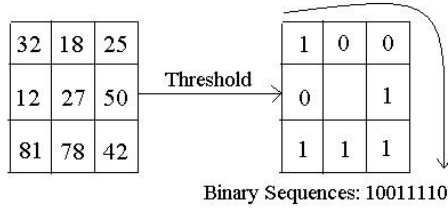

Fig. 1. *An example of LBP operator*

Multiresolution analysis can be achieved by choosing different values of m and R, where m denotes the number of neighboring pixels with respect to the center pixel, and R represents the distance from the center pixel to each of the neighboring pixels. Figure 2 illustrates circularly symmetric neighbor sets for different values of m and R.
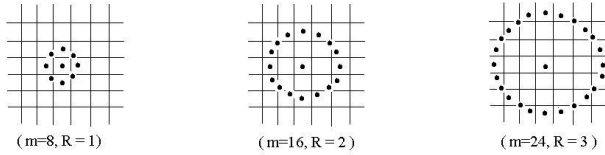

Fig. 2. *Circularly symmetric neighbor sets for different values of m and R.*

However, it is observed that in some facial expressions, the dominant patterns are not always uniform. The proportions of uniform patterns in some samples from the JAFFE database are shown in Table 1 with different values of m and R. Discarding the information of all non-uniform patterns may result in losing useful micro-pattern structural features in the face images.

Face Images	$m=8,R=1$	$m=16,R=2$	$m=24,R=3$
KA.FE2.46	62.04	52.84	43.25
KM.HA1.4	52.17	45.24	36.38
KM.HA4.7	48.72	35.81	28.92
YM.FE2.68	54.72	42.04	30.52

Table 1. *Proportions (%) of "uniform LBP" for some samples from the JAFFE database. It implies that the dominant patterns are not mainly the "uniform LBP" in some face images.*

In order to fully describe the dominant patterns contained in the face images, we extend the conventional LBP to ALBP. Each pattern

in the image is assigned a unique label by the following equation [7]

$$LBP(m, R) = \sum_{i=0}^{p-1} u(t_i - t_c)2^i. \qquad (1)$$

where t_c is the intensity of the center pixel, t_i is the intensity of the neighbor i, and $u(x)$ is the step function. It is clear that the LBP defined in Equation 1 is not rotation-invariant as the intensity value of t_i changes when the circle is rotated by a specific angle. Two patterns should be treated as the same type if one can be obtained from the other through rotating by a certain angle. Therefore, the ALBP pattern group is defined as:

$$ALBP^s(m, R) = \min(Cir(LBP(m, R), n)), \qquad (2)$$

where $n = 0, 1, ..., p-1$, $Cir(x, n)$ performs a circular anti-clockwise bitwise shift on the p-bit number by n times.

The histogram of the ALBP group is computed from the image and sorted in a descending order. Therefore, the first several entries of the histogram are guaranteed to be the dominant patterns from the image. Our experiments show that around 80% of the patterns in the image are sufficient for representing the information of the dominant patterns, and the 16 leading entries are enough to cover 80% of the whole ALBP histogram in the face image.

To further characterize the energy variation of the intensity magnitude in different local regions, we also apply the ALBP method to the gradient maps. The gradient of each pixel in the face image is defined as: $\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})^T$. We construct the ALBP histogram of the gradient magnitude map $|\nabla f|$ of the face image, following the same procedure of constructing the ALBP histogram in the intensity maps.

2.2. Decomposition of Face Images

In facial expression recognition, it is expected that some face regions (e.g. eyes, mouth) provide more important information than others. Therefore, each face image is partitioned into eight regions R_0, R_1, ..., R_7 as shown in Figure 3. Each region is assigned a weight. The larger a weight is, the more important is the region. The ALBP histograms for both intensity and gradient maps are extracted from each region. Then the feature vector is normalized to the range of $[-1, 1]$. Finally, the normalized feature vector is multiplied by its corresponding weight to obtain the region feature vector. As such, the region feature vector encodes the texture information in each local region. By concatenating all the region feature vectors together, global information of the entire face image can be obtained.
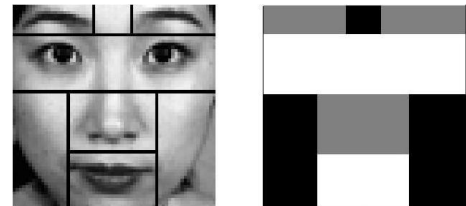

Fig. 3. *(a) An example of facial expression image divided into eight regions; (b) The weights set for different regions: black regions indicate weight 1.00, gray regions 2.00, and white regions 4.00.*

3. TSALLIS ENTROPY OF GABOR FILTERED RESPONSES

The ALBP features describe the high-frequency texture characteristics of the face image as they are computed by pixel-wise intensity differences. To effectively capture the low- and mid-frequency texture information, the Tsallis entropy is computed from the histogram of the Gabor filtered responses. The Tsallis entropy [9] is defined as:

$$S_q = \frac{1 - \sum_{j=0}^{m-1} h_j^q}{q - 1}, \tag{3}$$

where q is a continuous parameter, m denotes the maximum number of histogram bins, and h_j is the value of histogram entry i. The physical meaning of Tsallis entropy [9] is that it reflects the distribution characteristics of a histogram. Therefore, it can capture more detailed information with respect to the histogram distribution than the average magnitude response (i.e., mean value of the histogram) which was adopted by most traditional methods.

In this work, 16 ($= 4 \times 4$) Gabor filters are used corresponding to four different frequencies: $F_1 = 2.0$, $F_2 = 3.17$, $F_3 = 5.04$, and $F_4 = 8.0$, following the Gabor filter design strategy described in [13]. These filters slightly overlap each other with the Fourier domain evenly covered. Four different orientations are considered with respect to each frequency: $\theta_1 = 0°$, $\theta_2 = 45°$, $\theta_3 = 90°$, $\theta_4 = 135°$. Then for each filtered output, the Tsallis entropy is computed, and the 16 entropy values are used as the final features.

The Tsallis entropies capture low- and mid-frequency information of the face image as the corresponding Gabor filters' frequencies are fallen in such frequency range. Therefore, the Tsallis entropy features can complement the ALBP features well.

4. DISCRIMINATING APPEARANCE FEATURES

The extracted features above put more emphasis on low-level textural information of the face region rather than its global appearance. To better utilize the class information of the training data, we should also incorporate discriminating global appearance characteristics in the feature representation.

In typical appearance-based methods, $m \times n$ face images are often represented by points in an mn-dimensional space. However, coherent structure in the appearance of human faces leads to strong correlation between them, generating observations that lie on a low-dimensional subspace. The most widely used LDA approach [14] seeks an optimal projection $W_{opt} = \arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|}$ from the input space onto a lower-dimensional feature space. This maximizes the ratio of the between-class scatter to the within-class scatter. Here, we use an extended version of LDA, called null-space based LDA (NLDA) [10], to fully extract the discriminating appearance features for different facial expression categories. The number of NLDA features is set to $c - 1$, where c is the number of categories [14].

Up to now, we have collected three different types of features: ALBP in both intensity and gradient maps, Tsallis entropy of Gabor filtered responses, and null-space based LDA projection coefficients, which in turn capture the local intensity and energy variation patterns, the global low- and mid-frequency domain distributions, and the discriminating appearance characteristics of the face image. As human facial expressions are complex visual stimuli, we believe the combination of all three complementary features will lead to enhanced recognition performance.

5. EXPERIMENTAL RESULTS

The JAFFE database [11] is used to evaluate the facial expression recognition performance of the proposed method. In the JAFFE database, there are 10 persons. Each person has seven types of facial expressions: angry, disgust, fear, happy, neutral, sadness, and surprise (see Figure 4 for some examples in the database). There are three samples corresponding to each facial expression of each person. All the face images are resized to a fixed size of 64×64 pixels. Histogram equalization is performed to remove the illumination effects in the images. The support vector machine (SVM) [15] with an RBF kernel is used as the classifier in our work. In the experiments, two samples of each facial expression for each person are used to form the training set, and the remaining samples are used for testing. Each face image is partitioned into eight regions as shown in Figure 3. The proposed method is compared with two widely used approaches:

- **Uniform Local Binary Patterns** (**LBP**) [7]: The histogram of the uniform patterns in the face images, using $m = 8$ and $R = 2$ which is the same setting as in [6].
- **Average Magnitude of Gabor Filtered Responses** (**AMGFR**) [16]: The traditional Gabor filtered features, the average magnitude responses of the Gabor filtered outputs [16].

Fig. 4. *Some sample images from the JAFFE database*

5.1. Experiment on JAFFE Database with Original Resolution

In our first experiment, classification is performed using the original face images with resolution level 64×64 to evaluate the effectiveness of different approaches. Their classification accuracies are shown in Table 2, with the features of the proposed method listed in the last four rows: ALBP, Tsallis entropy (Tsallis), and NLDA in intensity images (NLDAI).

Features	Classification Accuracy %
AMGFR [16]	82.46
LBP [7]	85.57
ALBP	88.26
Tsallis	85.36
ALBP + Tsallis	91.89
ALBP + Tsallis + NLDAI	94.59

Table 2. *Performance comparison of different approaches with resolution level 64×64 for the images from the JAFFE database*

The classification accuracies summarized in Table 2 show that the Tsallis entropy outperforms the average magnitude response fea-

tures computed from the Gabor filters. This is because the Tsallis entropy describes the whole histogram distribution information of the filtered images, while the average magnitude response only computes the mean value of the histogram. The ALBP features outperform the conventional LBP as the ALBP histogram can always represent the most essential and dominant patterns in the face images. Moreover, the ALBP and Tsallis entropy features complement well with each other as discussed in Section 3, which is empirically reflected by the high classification accuracy of 91.89%. The dimensionality-reduced projections of the intensity image onto a low-dimensional subspace using NLDA produce the discriminating features of the global facial appearance. Therefore, the NLDA features (i.e., global appearance features) can further complement the ALBP and Tsallis entropy features (i.e., texture features), which is also reflected by the highest classification accuracy of 94.59% in the experiment.

5.2. Experiment on JAFFE Database with Lower Resolution

In many real-world applications, the resolution of face images is usually low due to various factors such as surrounding environment and imaging equipment. Therefore, it is important to evaluate the recognition performance of the proposed approach also under low-resolution conditions. In the second experiment, the resolution level of the original images (64×64) is downsampled to 48×48, 32×32 and 16×16, respectively, to observe the relationship between the classification accuracy and the resolution level of the face images. The classification accuracies for different methods are listed in Table 3.

Features	Classification accuracy (%)		
	48×48	32×32	16×16
AMGFR [16]	78.13	67.83	56.35
LBP [7]	81.44	77.28	68.02
ALBP	84.27	82.74	75.39
Tsallis	79.25	71.04	63.81
ALBP + Tsallis	87.31	85.73	80.40
ALBP + Tsallis + NLDAI	90.54	88.82	84.62

Table 3. *Performance comparison of different approaches with resolution levels 48×48, 32×32 and 16×16 for the images from the JAFFE database*

It is observed that the ALBP approach is more effective and stable than the conventional LBP when the resolution of the input face images is low. The Tsallis entropy features still outperform the traditional Gabor filtered features under all three low resolution levels. Combining the ALBP and Tsallis entropy features produces acceptable classification accuracy (80.40%) even at the lowest resolution level (16×16). The proposed method listed in the last row gives the highest classification accuracies under all three resolution levels.

6. CONCLUSION

This paper describes a novel approach to the automatic recognition of facial expressions. Motivated by our recent exploration of effective image features for the task of texture recognition and face recog-

nition, we combine advanced LBP and Tsallis entropy information with discriminating appearance features to characterize seven facial expression styles. Experimental results based on the JAFFE database demonstrate the effectiveness and flexibility of our proposed method when compared with two widely used methods in the literature.

7. REFERENCES

[1] Y. Tian, T. Kanade, and J.F. Cohn, *Facial Expression Analysis*, Springer, 2003.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71–86, January 1991.

[3] Y. Tian, "Evaluation of face resolution for expression analysis," in *Workshop on Face Processing in Video*. IEEE, 2004.

[4] L. Ma, Y. Xiao, K. Khorasani, and R. Ward, "A new facial expression recognition technique using 2-d dct and k-means algorithm," in *Proc. of ICIP*, 2004, pp. 1269–1272.

[5] Z. Zhang, M. Lyons, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perception," in *IEEE FG*, April 1998.

[6] C. Shan, S. Gong, and P.W. McOwan, "Robust facial expression recognition using local binary patterns," in *ICIP*. IEEE, 2005, pp. 370–373.

[7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, July 2002.

[8] S.Liao, W.K. Law, and A.C.S. Chung, "Combining microscopic and macroscopic information for rotation and histogram equalization invariant texture classification," in *ACCV (to appear)*, 2006.

[9] C. Tsallis, "Nonextensive statistics: Theoretical, experimental and computational evidences and connections," *Brazilian Journal of Physics*, vol. 29, pp. 1–35, 1999.

[10] R. Huang, Q. Liu, H. Lu, and S. Ma, "Solving the small size problem of lda," in *Proc. of ICPR*, 2002, vol. 3, pp. 29–32.

[11] M. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. on PAMI*, vol. 21, pp. 1357–1362, 1999.

[12] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," in *ECCV*, 2004, pp. 469–481.

[13] J.F. Camapum Wanderley, *Color-based recognition for remote sensing in environmental systems*, Ph.D. dissertation, Coventry Univ., Coventry, U.K., February 1999.

[14] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. on PAMI*, vol. 19, pp. 711–720, July 1997.

[15] V.N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[16] A.C. Bovik, M. Clark, and W.S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Transactions on PAMI*, vol. 12, pp. 55–73, Jan 1990.