LOCAL FEATURE EXTRACTION METHODS FOR FACIAL EXPRESSION RECOGNITION

Seyed Mehdi Lajevardi, Zahir M. Hussain

School of Electrical & Computer Engineering, RMIT University
City Campus, Swanston St., Melbourne, Australia
seyed.lajevardi@rmit.edu.au, zmhussain@ieee.org
http://www.rmit.edu.au

ABSTRACT

In this paper we investigate the performance of different feature extraction methods for facial expression recognition based on the higher-order local autocorrelation (HLAC) coefficients and local binary pattern (LBP) operator. Autocorrelation coefficients are computationally inexpensive, inherently shift-invariant and quite robust against changes in facial expression. The focus is on the difficult problem of recognizing an expression in different resolutions. Results indicate that LBP coefficients have surprisingly high information content.

1. INTRODUCTION

Facial expression is a visible manifestation of the affective state, cognitive activity, intention, personality, and psychopathology of a person; it not only expresses our emotions, but also provides important communicative cues during social interaction. Reported by psychologists, facial expression constitutes 55% of the effect of a communicated message while language and voice constitute 7% and 38% respectively. So it is obvious that analysis and automatic recognition of facial expression can improve human-computer interaction or even social interaction.

An automatic classification of facial expressions consists of two stages: feature extraction and feature classification. The feature extraction is a key importance to the whole classification process. If inadequate features are used, even the best classifier could fail to achieve accurate recognition. In most cases of facial expression classification, the process of feature extraction yields a prohibitively large number of features and subsequently a smaller sub-set of features needs to be selected according to some optimality criteria.

Local Binary Patterns (LBP), first proposed by Ojala et al. [5], is a powerful means of texture description. The block-based approach based on local binary patterns is extended for facial expression recognition.

Higher-Order Local Auto-Correlation (HLAC) features is proposed by Otusu [3] for Feature Extraction. HLAC features, an extension of autocorrelation features (second-order statistics), are based on higher-order statistics (HOS), however the dimensionality of the resulting data is so high.

The dimensionality reduction can be achieved by selection more informative features based on mutual information [8], [9], [10]. In this study, the minimum redundancy - maximum relevance (MRMR) [6] was investigated to select the optimum features for classification. The MRMR algorithm is based on mutual information. The mutual information was used as an objective criterion in selection of optimal subsets of features in a feature reduction task. In contrast to the classical correlation-based feature selection methods, the mutual information can measure arbitrary relations between variables and it does not depend on transformations applied to different variables. It can be potentially useful in problems where methods based on linear relations between data are not performing well.

A functional block-diagram of the proposed facial expression recognition system is illustrated in Fig.1.

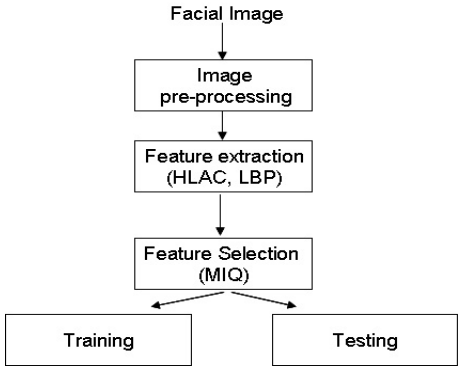

Figure 1: Block diagram of the facial recognition system.

The remainder of this paper describes the methods, experiments and results. Section 2 describes the database used to train and test the facial expression recognition system. Section 3 explains the image pre-processing steps. In section 4, the feature extraction is explained. Section 5 describes the feature selection. Section 6 explain the classification. Section 7 contains experimental results and in section 8 final conclusions are presented.

2. IMAGE DATA

The images selected from the Cohn-Kanade database [1] were used to train and test the facial expression recognition system. The Cohn-Kanade database consists of approximately 500 image sequences from 100 subjects. The subjects range in age from 18 to 30 years. Sixty-five percent of subjects are female; fifteen percent are African-American and three percent Asian or Latino. The total of 359 static images was selected for the purpose of this study. The selected images represent 100 different subjects expressing all or some of the six emotions: anger, disgust, fear, happiness, sadness and surprise. For each subject only one image per expression was used. Some subjects did not have images corresponding to all of the six expressions. Fig.2 shows expression samples from the Cohn-Kanade database.

Figure 2: Examples of original images from the Cohn-Kanade database.

3. IMAGE PRE-PROCESSING

The image pre-processing procedure is a very important step in the facial expression recognition task. The aim of the pre-processing phase is to obtain sequences of images which have normalized intensity, are uniform in size and shape, and depict only the face region. The image intensity was normalized using the histogram equalization. The face area of an image was detected using the Viola-Jones method [2] based on the Haar-like features and the AdaBoost learning algorithm.

The Viola and Jones method is an object detection algorithm providing competitive object detection rates in real-time. It was primarily designed for the problem of face detection. The features used by Viola and Jones are derived from pixels selected from rectangular areas imposed over the picture and show high sensitivity to the vertical and horizontal lines.

AdaBoost, is an adaptive learning algorithm that can be used in conjunction with many other learning algorithms to improve their performance. AdaBoost is adaptive in the sense that subsequent classifiers built iteratively are made to fix instances misclassified by previous classifiers. At each iteration a distribution of weights is updated such that, the weights of each incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on those examples.

The final stage of the pre-processing involved detection of an image which depicted certain emotion with the maximum level of arousal (emotion intensity). This was done using the minimum mutual information (MI) criterion. For each frame, the mutual information between the current frame and the initial frame was calculated, and the frame with the minimum mutual information was selected as the frame that represents an emotion with the maximum arousal [7], [8].

Figure 3: Six facial expression images after pre-processing.

4. FEATURE EXTRACTION

The feature extraction phase represents a key component of any pattern recognition system. In this study, a local approach, which extracts features from the picture locally, was implemented.

4.1 Higher-order local Auto Correlation

The features are generated using higher-order local autocorrelation. The Nth-order autocorrelation functions, extensions of autocorrelation functions, are defined as:

$$x(a_1, a_2, ..., a_N) = \int f(r) f(r + a_1) f(r + a_N) dr \qquad (1)$$

where $f(r)$ denotes the intensity at the observing pixel r,and $a_1, a_2, \cdots, a_N$ are N displacements. HLAC features [3] are primitive image features based on Eq. (1). Their orders and displacements are arbitrary. However, higher order features with a large displacement region become extremely numerous. Hence, the original HLAC features are restricted up to the second order (three-point relations) and within a 3×3 displacement region. They are represented by 25 mask patterns with 0, 1 and 2 displacements (25 mask patterns in Fig. 4). Each mask pattern is scanned over the entire image, and for each possible position, the product of the pixels marked in white is computed.

All the products corresponding to a mask are then summed so as to provide one feature. This operation is performed using 25 different mask pattern to create the feature vector for each facial image. Each feature value represents the power spectrum of the mask pattern, which corresponds to a basis functions of frequency analysis [4]. Roughly comparison with a Fourier transform, the mask size corresponds to the frequency component, and the distribution of the displacements corresponds to the direction component. Since the HLAC features use the information of two-dimensional distributions as well as the directions, they analyze an image more closely.

Furthermore, we use large mask patterns to support large displacement regions (Fig.5) and extract the features of low resolutions or low frequencies. Therefore, we use masks of different sizes together and construct multi-resolution features.

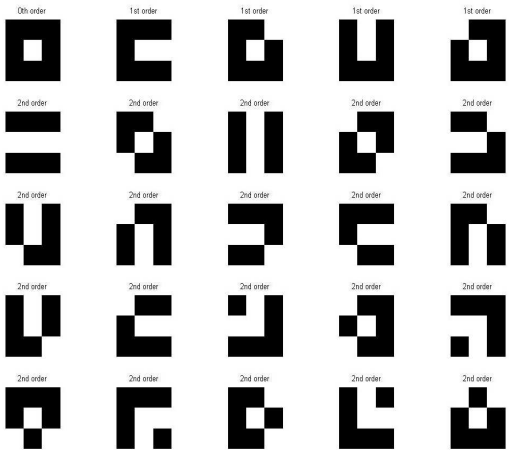

Figure 4: 25 mask patterns of the HLAC features (3x3).

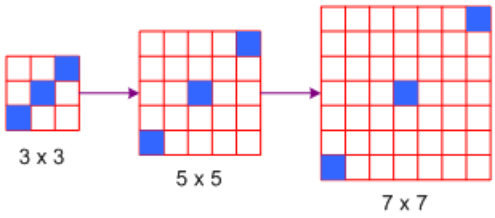

Figure 5: An extention of HLAC features.

4.2 Local Binary Pattern Operator

The original LBP operator, introduced by Ojala [5], is a powerful method of texture description. The original 3×3 neighborhood is thresholded by the value of the center pixel. The values of the pixels in the thresholded neighborhood are multiplied by the binomial weights given to the corresponding pixels. Finally, the values of the eight pixels are summed to obtain the LBP number for this neighborhood. An illustration of the basic LBP operator is shown in Fig.6.

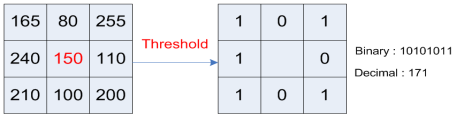

Figure 6: Illustration of the basic LBP operator.

An extension to the original operator is to use so called uniform patterns. A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. Ojala noticed that in their experiments with texture images, uniform patterns account for a bit less than 90% of all patterns when using the (8,1) neighborhood and for 70% in (16,2) neighborhood. We use the following notation for the LBP operator: $LBP^{u2}_{P,R}$ means using the operator in a neighborhood of P sampling points on a circle of radius R. Superscript ^{u2} stands for using uniform patterns and labeling all remaining patterns with a single label. In this work, $LBP^{u2}_{8,2}$ is applied to extract LBP code for each pixel of face images, generating LBP faces. All feature values are quantified into 59 bins according to uniform strategy. A histogram of the labeled image $f_l(x,y)$ can be defined as:

$$H_i = \sum_{x,y} I\{f_l(x,y) = i\}, i = 0,...,n-1 \tag{2}$$

in which n is the number of different labels produced by the LBP (in this work, LBP coefficients are quantified into 59 bins, so n is 59) and

$$I\{A\} = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false} \end{cases} \tag{3}$$

This LBP histogram contains information about the distribution of the local micro-patterns, such as edges, spots, and flat areas over the whole image.

Face images can be seen as a composition of micropatterns which can be effectively described by the LBP features. However, a LBP histogram computed over the whole face image encodes only the occurrences of the micropatterns without any indication about their locations. Considering shape information of faces, face images are divided into small regions $R_0, R_1, \cdots, R_{m-1}$ to extract LBP features (See Fig.7 for an illustration). In this study, we used different size of subregions for different image resolutions. The LBP features extracted from each sub-region are concatenated into a single, spatially enhanced feature histogram defined as:

$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \in R_j\} \tag{4}$$

where $i = 0, \cdots, n-1, j = 0, \cdots, m-1$

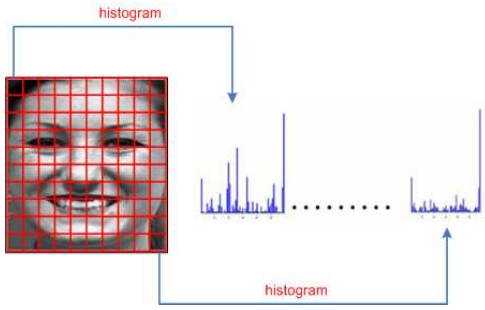

Figure 7: A facial image is divided into 100 small regions from which LBP histograms are extracted and concatenated into a single histogram.

5. FEATURE SELECTION

5.1 Minimum Redundancy - Maximum Relevance Criteria

A feature selection method based on the mutual information quotient (MIQ) [6] criterion was investigated. If a feature vector has expressions randomly or uniformly distributed in different classes, its mutual information with these classes is zero. If a feature vector is strongly differentially expressed for different classes, it should have large mutual information. Thus we use mutual information as a measure of relevance of feature vectors. the mutual information I of two variables x and y is defined based on their joint probabilistic distribution $p(x,y)$ and the respective marginal probabilities $p(x)$ and $p(y)$:

$$I(x;y) = \sum_{i,j} p(x_i,y_j) \log \frac{p(x_i,y_j)}{p(x_i)p(y_j)} \tag{5}$$

The idea of minimum redundancy is to select the feature vectors such that they are mutually maximally dissimilar. Minimal redundancy will make the feature set a better representation of the entire data set. Let S denote the subset of features that we are seeking. The minimum redundancy condition is

$$\min W_I, \quad W_I = \frac{1}{|S|^2} \sum_{f_i,f_j \in S} I(f_i,f_j) \tag{6}$$

where $I(f_i,f_j)$ is the mutual information between f_i and f_j, and $|S|$ is the number of features in S.

To measure the level of discriminant powers of features when they are differentially expressed for different targeted

classes, we again use mutual information $I(C,f_i)$ between targeted classes $C=\{C_1,C_2,,C_6\}$. Thus $I(C,f_i)$ quantifies the relevance of f_i for the classification task. Thus the maximum relevance condition is to maximize the total relevance of all features in S:

$$\max V_I, \quad V_I = \frac{1}{|S|}\sum_{f_i \in S} I(C,f_i) \tag{7}$$

The minimum redundancy, maximum relevance feature set is obtained by optimizing the conditions in Eqs.(6) and (7) simultaneously. Optimization of these two conditions requires combining them into a single criterion function as follow:

$$\max(V_I/W_I) \tag{8}$$

In this algorithm, the first feature is selected according to Eq. (7), i.e. the feature with the highest $I(C,f_i)$. The rest of features are selected in an incremental way: earlier selected features remain in the feature set. Suppose we already select m features for the set S, we want to select additional features from the set $F_S = F_T - S$. We optimize the following two conditions:

$$\max I(C,f_i), \quad f_i \in F_S \tag{9}$$

$$\min \frac{1}{|S|}\sum_{f_j \in S} I(f_i,f_j), \quad f_i \in F_S \tag{10}$$

By combining Eq.(6),(7) and (8) we have the following equation to calculate the MIQ for feature selection:

$$\max\{\frac{I(f_i,C)}{\frac{1}{|S|}\sum I(f_i,f_j)}\}, \quad f_i \in F_S, f_j \in S \tag{11}$$

6. CLASSIFICATION

The facial expressions depicted in images were classified using Naive Bayesian (NB) classifier. The NB classifier is a probabilistic method that has been shown to be effective in many classification problems. It assumes that the presence (or lack) of a particular feature of a class is unrelated to the presence (or lack) of any other feature. The classification decision is made using the following formula:

$$C = \arg\max\{P(C_i)\prod P(f_i|C_i)\} \tag{12}$$

where $P(f_i|C_i)$ are conditional tables (or conditional density) learned in training using examples. Despite the independence assumption, NB has been shown to have very good classification performance for many real data sets on par with many more sophisticated classifiers.

7. EXPERIMENTAL RESULTS

Facial expression image sequences from the Cohn- Kanade database [1] were used to train and test the facial expression recognition system.The Cohn-Kanade database consists of approximately 400 image sequences from 100 subjects. The subjects range in age from 18 to 30 years. Sixty-five percent of subjects are female; fifteen percent were African-American and three percent Asian or Latino. Each sequence contained 12-16 frames. The image sequences expressing different stages of an expression development, starting from a low arousal stage, reaching a peak of arousal and then declining. The facial expressions of each subject represented six basic emotions: anger, disgust, fear, happiness, sadness and surprise.

The training set size was 216 and the test size was 172 image sequences. Each test was performed 3 times using randomly selected testing and training sets and an average result was calculated. The training set contained the same number of image sequences for each expression. The subjects represented in the training set were not included in the testing set of images, thus ensuring a person-independent classification of facial expressions.

Two different approaches to the facial expression recognition task were compared. In the first approach HLAC features were used to extract the features from the images. In the second approach the features were generated based on LBP operator. Then we used MIQ algorithm to select the optimum subset for classification. The best subset of feature that we selected for classification is S=55. In all cases, the test images were classified using the NB classifier. The results are shown for different resolutions from 16×16 to 128×128. Table.1 and Table.2 shows the confusions for different expression in low resolution (16x16). Fig.8 illustrated the average correct classification for both HLAC and LBP operator. It is shown that the classification result based on LBP features for low resolution samples are better than HLAC features,however, for high resolution samples the classification performance for LBP features is more accurate than HLAC features.

Table 1: Confusion table for HLAC features

	A	D	F	H	S	Su
A	56.2	25.6	0	0	18.2	0
D	22.2	55.6	14.8	3.7	3.7	0
F	5.6	6.9	52.8	27.8	0	6.9
H	0.6	4.5	19.2	75.0	0	0.7
S	16.1	6.1	5.1	0	72.7	0
Su	0	12.5	4.2	0	0	83.3
Average	**65.5**					

A: anger D: disgust F: fear H: happy S: sad Su: surprise

Table 2: Confusion table for LBP operator

	A	D	F	H	S	Su
A	58.4	19.0	4.8	4.8	13.0	0
D	11.1	57.4	0	20.3	11.1	0
F	11.1	5.6	58.1	19.6	3.7	1.9
H	2.3	5.1	15.2	75.2	0	2.2
S	4.8	12.0	3.7	1.2	77.0	1.2
Su	0	1.6	5.3	2.4	2.8	88.0
Average	**69.0**					

A: anger D: disgust F: fear H: happy S: sad Su:surprise

8. CONCLUSION

A comparison of feature extraction methods for the facial expression recognition from image sequences was presented and tested. The method is fully automatic and includes: face detection, maximum arousal detection, feature extraction, selection of optimal features and classification. The lo-

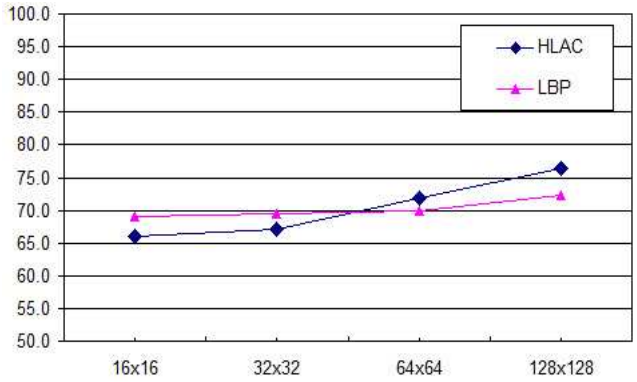

Figure 8: A comparision of recognition rates between HLAC features and LBP operator.

cal binary pattern operator (LBP) in low resolution images increased the average percentage of correct classifications from 65.9% to 69% and from 67.2% to 69.5% for 16×16 and 32×32 respectively.

The presented feature selection method is based on the mutual information(MIQ) criterion, and does not assume linear dependencies between data. It can therefore handle arbitrary relations between the pattern coordinates and the different classes. The additional advantages of the feature selection based on the MI criterion include computational simplicity and invariance to the data transformations. The system not only offers an optimized feature selection, but also automatically finds an optimal frame to represent a given class of emotion.

An overall improvement of the classification results and the discrimination between different facial expressions was observed when using LBP operator. The accuracy for high resolution images based on HLAC features was better than LBP features, however the complexity and time consuming were more discriminate in HLAC feature extraction process. Though the LBP operator can improve the total recognition ratios, its not as good as HLAC features in high resolution images. We will experiment more to find the reason and combine other method to solve these problems in future work.

REFERENCES

[1] Kanade, T., Cohn, J. F., and Tian, Y.: "Comprehensive database for facial expression analysis," *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France,pp. 46-53, 2000.

[2] Viola P., Jones M., "Robust Real-time Object Detection," *International Journal of Computer Vision*, 2004.

[3] Otsu, N., and Kurita., T.: "A new scheme for practical flexible and intelligent vision systems," *In Proceedings of the IAPR Workshop on Computer Vision*, pp. 431-435, 1988.

[4] Toyoda, T., and Hasegawa., O.: "Texture classification using extended higher order local autocorrelation features," *Proceedings of the 4th International Workshop on Texture Analysis and Synthesis*, pp. 131-136, 2005.

[5] Ojala, T., Pietikainen, M., and Harwood, D.: "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, Vol 29, No. 1, pp. 51-59, 1996.

[6] Peng, H., Long, F., and Ding, C.: "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 8, pp.1226-1238, 2005.

[7] Lajevardi, S. M., Lech, M., "Facial Expression Recognition from Image Sequences Using Optimised Feature Selection," *IVCNZ08*, Christchurch, New Zealand, 2008.

[8] Lajevardi, S. M., Hussain, Z. M., "Facial Expression Recognition: Gabor Filters versus Higher-Order Correlators," *ICCCP09*, Muscat, Oman, 2009.

[9] Lajevardi, S. M., Hussain, Z. M., "Facial Expression Recognition Using Log-Gabor Filters and Local Binary Pattern Operators," *ICCCP09*, Muscat, Oman, 2009.

[10] Lajevardi, S. M., Hussain, Z. M., "Feature Selection for Facial Expression Recognition Based on Mutual Information," *IEEEGCC09*, Kuwait, 2009.