Facial Expression Recognition Using Log-Gabor Filters and Local Binary Pattern Operators

Seyed Mehdi Lajevardi, Zahir M. Hussain
School of Electrical & Computer Engineering, RMIT University, Melbourne, Australia
seyed.lajevardi@rmit.edu.au, zmhussain@ieee.org

***Abstract*—This study investigates two different methods of feature extraction for person-independent facial expression recognition from images. The logarithmic Gabor filters and the local binary pattern operator (LBP) were used for feature extraction. Then, the optimum features were selected based on minimum redundancy and maximum relevance algorithm (MRMR). Six different facial expressions were considered. The selected features were classified using the naive bayesian (NB) classifier. The percentage of correct classifications varied across different expressions from 62.8% to 90.5% for the log-Gabor filter approach,and from 71.8% to 94.0% for the LBP approach.**

Experiments carried out on Cohn-Kanade database showed comparable performance between Log-Gabor filters and LBP operator, with a classification accuracy of around 82.3% and 81.7% respectively. This was achieved on low-resolution images, without the need to precisely locate facial points on each face image.

-Index Terms- Feature selection, minimum redundancy maximum dependency, mutual information quotient, log-Gabor filters, LBP operator, Facial Expression Recognition.

I. Introduction

FACIAL expression recognition plays important role in a variety of applications such as automated tools for behavioral research, bimodal speech processing, video conference, airport security and access control, building (embassy) surveillance/monitoring, human computer intelligent interaction and perceptual interfaces, etc.

Various methods for recognizing human facial expressions from face images have been proposed and their performance has been evaluated with databases of face images with variations in expressions. Generally, there are two categories of feature representation: geometric features and appearance feature. Appearance features have been demonstrated to be better than geometric features, because geometric features are very sensitive to noises, especially illumination noise.

Local Binary Patterns (LBP), first proposed by Ojala et al. [2], is a powerful means of texture description. The block-based approach based on local binary patterns is extended for facial expression recognition.

Gabor wavelets are reasonable models of visual processing in primary visual cortex and are one of the most successful approaches to describe local appearance of the human face [1]. However they have two main limitations. The maximum bandwidth of a Gabor filter is limited to approximately one octave and Gabor filters are not optimal if one is seeking broad spectral information with maximal spatial localization. The logarithmic Gabor filters proposed by Field [3] are known to provide excellent simultaneous localization of spatial and frequency information, however, the dimensionality of the resulting data is high. For this reason, an optimal sub-set of the log-Gabor filter data was selected.

Selecting a small subset of features out of the thousands of features is important for accurate classification. instead of using all available variables (features or attributes) in the data, one selectively chooses a subset of features to be used in the discriminant system. There are a number of advantages of feature selections: (1) dimension reduction to reduce the computational cost; (2) reduction of noises to improve the classification accuracy; (3) more interpretable features or characteristics that can help classify the expression more accurately[7].

Feature selection methods that are adequate for simple distributions of patterns belonging to different classes fail in classification tasks with more complex distributions and overlapping boundaries. Methods such as correlation assume linear dependencies between data, and cannot handle arbitrary relations between the pattern coordinates and the different classes. Commonly used data reduction techniques such as the principal component analysis (PCA) [10], are not invariant under linear transformations such as data scaling used in the pre-processing stage.

In this study, the minimum redundancy - maximum relevance (MRMR)[7] 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 sub-sets 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.

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 method based on the log-Gabor Filter and LBP operator is explained. Section 5 describes the feature selection. Section 6 contains the experimental results and in Section 7 final conclusions are presented.

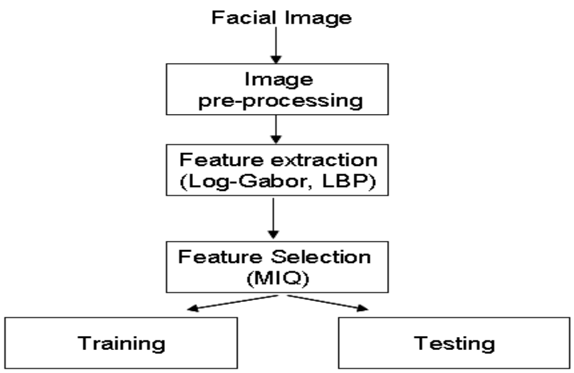

Fig. 1: Block diagram of the facial recognition system.

II. IMAGE DATA

Static expression images selected from the Cohn-Kanade database [6] 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.

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

III. IMAGE PRE-PROCESSING

The aim of the pre-processing phase was to obtain images which have normalized intensity, uniform size and shape, and depict only a face expressing certain emotion. As shown in the examples in Fig.2, the image backgrounds were plain without any textile. The parts of images that contained only the faces were extracted using the Sobel operator [11]. After the Sobel kernels were applied, the area of the face was found based on the blob analysis [9]. In image processing, a blob is defined as a region of connected pixels. The blob analysis algorithm identifies these regions in an image, and places them in one of two categories: the foreground (typically pixels with a non-zero value) or the background (pixels with a zero value). The parts representing faces were cut out from the images and their histograms were equalized. Finally, the images were scaled to the same size of 60×60 pixels. Fig.3 shows examples of images after the pre-processing.

Fig. 3: Examples of images after the pre-processing step.

IV. FEATURE EXTRACTION

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

A. Log Gabor Filters

Gabor filters are commonly recognized [8], [11], [12] as one of the best choices for obtaining localized frequency information. They offer the best simultaneous localization of spatial and frequency information. However, one cannot construct Gabor functions of arbitrarily wide bandwidth and still maintain a reasonably small DC component in the even-symmetric filter. Alternatively, we turn to choose the Log-Gabor filter as the filtering kernel [13]. There are two important characteristics to note. Firstly, Log-Gabor functions, by definition, always have no DC component, which contributes to improve the contrast ridges and edges of images. Secondly, the transfer function of the Log-Gabor function has an extended tail at the high frequency end, which enables us to obtain wide spectral information with localized spatial extent and consequently helps to preserve true ridge structures of images. Gabor functions have Gaussian transfer functions when viewed on the linear frequency scale. Similarly, Log-Gabor functions have Gaussian transfer functions when viewed on the logarithmic frequency scale. The log-Gabor filters are defined in the frequency domain using polar coordinates by the transfer function $H(f, \theta)$ constructed as a following product:

$$H(f, \theta) = H_f \times H_\theta \tag{1}$$

the radial component H_f controlling the bandwidth that the filter responds to, and the angular component H_θ, controlling the spatial orientation that the filter responds to. The 2D log-Gabor filters can be represented in a polar form as:

$$H(f, \theta) = \exp\{\frac{-[\ln(\frac{f}{f_0})]^2}{2[\ln(\frac{\sigma_f}{f_0})]^2}\} \exp\{\frac{-(\theta - \theta_0)^2}{2\sigma_\theta^2}\} \tag{2}$$

where f_0 is the filter's center frequency, and θ_0 the filter's direction. The constant σ_f, defines the radial bandwidth B in octaves:

$$B = 2\sqrt{\frac{2}{\ln 2}} \times |\ln(\frac{\sigma_f}{f_0})| \tag{3}$$

The constant σ_θ, defines the angular bandwidth $\Delta\Omega$ in radians:

$$\Delta\Omega = 2\sigma_\theta\sqrt{\frac{2}{\ln 2}} \tag{4}$$

In the study described here, the ratio σ_f/f_0 was kept constant for varying f_0, B was set to one octave and the angular bandwidth was set to $\Delta\Omega = \Pi/4$ radians. This left only σ_f, to be determined for a varying value of f_0. Five scales and eight orientations were implemented to extract features from face images. This lead to 40 filter transfer functions $\{H_1, H_2, ..., H_{40}\}$ representing different scales and orientations.

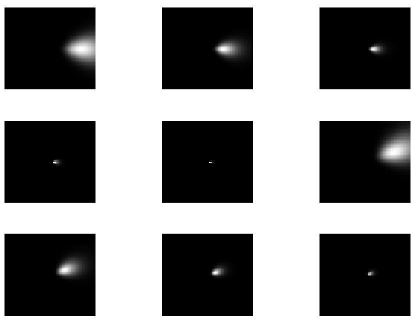

Fig. 4: Sample of Log-Gabor filters in frequency domain.

B. Local Binary Pattern Operator

The original LBP operator, introduced by Ojala [2], 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.5.

Fig. 5: Illustration of the basic LBP operator.

An extension to the original operator is to use so called uniform patterns [4]. 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 [5]. 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 labelled image $f_l(x, y)$ can be defined as:

$$H_i = \sum_{x,y} I\{f_l(x,y) = i\}, i = 0, ..., n-1 \tag{5}$$

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

$$I\{A\} = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false} \end{cases} \tag{6}$$

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.6 for an illustration). The LBP features extracted from each sub-region are concatenated into a single, spatially enhanced feature histogram defined as:

$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \in R_j\} \tag{7}$$

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

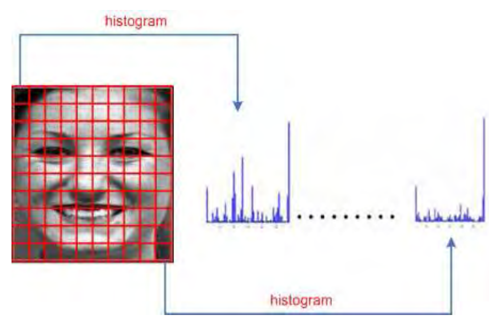

Fig. 6: A facial image is divided into 100 small regions from which LBP histograms are extracted and concatenated into a single histogram.

V. FEATURE SELECTION

A. Minimum Redundancy - Maximum Relevance Criteria

A feature selection method based on the mutual information quotient (MIQ) [7] 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 $p(y)$:

$$I(x;y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{8}$$

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

$$\min W_I, \quad W_I = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i, f_j) \tag{9}$$

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 genes 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:

$$\max V_I, \quad V_I = \frac{1}{|S|} \sum_{f_i \in S} I(C, f_i) \tag{10}$$

The minimum redundancy, maximum relevance feature set is obtained by optimizing the conditions in Eqs.(9) and (10) simultaneously. Optimization of these two conditions requires combining them into a single criterion function as follow:

$$\max(V_I / W_I) \tag{11}$$

In this algorithm, the first feature is selected according to Eq. (10), 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:

$$\max I(C, f_i), \quad f_i \in F_S \tag{12}$$

$$\min \frac{1}{|S|} \sum_{f_j \in S} I(f_i, f_j), \quad f_i \in F_S \tag{13}$$

By combining Eq.(9),(10) and (11) we have the following equation to calculate the MIQ for feature selection:

$$\max\{\frac{I(f_i, C)}{\frac{1}{|S|}\sum I(f_i, f_j)}\}, \quad f_i \in F_S, f_j \in S \tag{14}$$

VI. Experimental Results

Facial expression recognition tests were performed using static images from the Cohn-Kanade dataset. A total of 388 face images from 100 subjects were selected. The images were depicting six different facial expressions: anger, disgust, fear, happiness, sadness and surprise. In the training phase 180 images were used and in the testing phase 208 images were classified. The images used in the testing set were not included in the training set. 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. Because of the limited number of samples,each test was performed 3 times using randomly selected testing and training sets and an average results were calculated.

Two different approaches to the facial expression recognition task were compared. In the first approach 40 log-Gabor filters 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. In all cases, the test images were classified using the NB classifier. Fig.7 illustrate the accuracy for different feature number for both Log-Gabor filters and LBP operator.

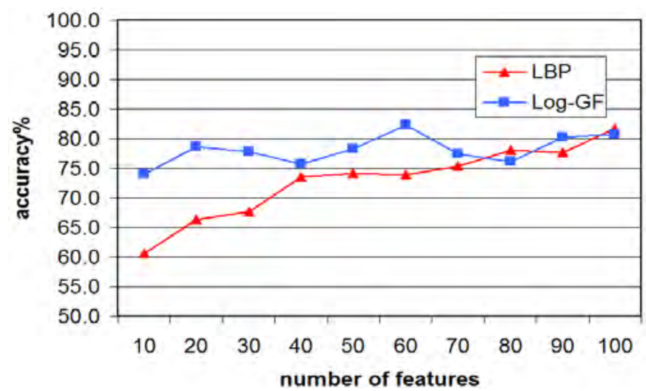

Fig. 7: The correct classification results for different number of features.

On average, the log-Gabor filters classified correctly 82.3% of cases, whereas the LBP operator method gave the overall rate of 81.7%. The percentage of correct classifications varied across different facial expressions. Table.I and Table.II shows the classification rate for different expressions. Fig.8 shows comparison of correct classification results between two methods.

Table.III shows an example of the CPU times corresponding to different stages of the classification process. The times are given (CPU speed 2.4 GHz and 2GB RAM) for two cases: the classification process for log-Gabor filters and the classification process for LBP operator. From Table.III we can see that the feature extraction based on log-Gabor filters is time consuming.

VII. Conclusion

In this paper, we studied two different methods of feature extraction for facial expression recognition system: the log-Gabor filters and LBP operator. The results showed that the log-Gabor method outperformed the LBP method producing the largest improvement in the classification accuracy and in the discrimination between different facial expressions. The LBP operator needs less memory and has less features (5900 features for each sample) than log-Gabor filters (36000 features for each sample).

For feature selection we used the MRMR method based on Mutual information quotient. The MIQ algorithm does not assume linear dependencies between data, and can 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.

Though the log-Gabor filters method have better recognition ratios than LBP operator, its not good to some expressions and also the computational load is complex and time consuming. In future works we will experiment more to find the reason and combine other methods to solve these problems.

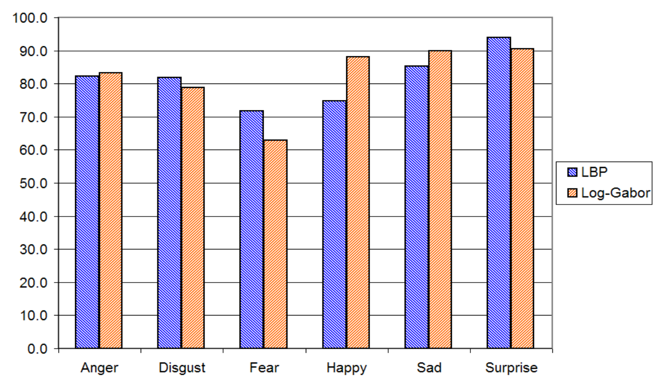

Fig. 8: A Comparison of correct classification for six facial expressions.

TABLE I: Percentage of correct classifications using LBP operator.

	A	D	F	H	S	Su
A	82.2	6.7	0	3.3	4.5	3.3
D	3.0	81.8	6.1	3.0	6.1	0
F	0	0	71.8	23.1	1.3	3.8
H	2.6	1.5	18.5	74.9	2.1	0.5
S	2.2	5.8	2.5	0	85.3	4.2
Su	1.8	1.2	0.6	0	2.4	94.0
Average	**81.7**					

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

TABLE II: Percentage of correct classification using log-Gabor filters.

	A	D	F	H	S	Su
A	83.3	3.3	3.3	0	10.0	0
D	12.1	78.8	3.0	3.0	3.0	0
F	5.1	0	62.8	17.9	11.5	2.6
H	3.6	0	7.2	88.2	0.5	0.5
S	8.3	0	0	0	90.0	1.7
Su	5.4	1.2	0	0	3.0	90.5
Average	**82.3**					

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

TABLE III: Comparison of the CPU time (sec).

	Log-Gabor	LBP Operator
Pre-Processing	1.05	1.05
Feature Extraction	10.70	0.17
Classification	0.96	0.49
Entire Process	12.71	1.71

REFERENCES

[1] Lee T. S., "Image Representation Using 2D Gabor Wavelets", IEEE Trans. *Pattern Analysis and Machine Intelligence*, Vol 18, No. 10, pp.959-971, 1996.

[2] 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.5159, 1996.

[3] Field D.J. statistics of natural, "Relations between the images and the response properties of cortical cells",*Jour. of the Optical Society of America*, pp. 2379-2394, 1987.

[4] Ojala T., Pietikainen M., and Maenpaa M., Multiresolution gray-scale and rotation invariant texture classification width local binary patterns,*IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 24, No. 7, pp.971987, 2002.

[5] QUAN ZHAO, BAO-CHANG P., JIAN-JIA P., YUAN-YAN T.,"Facial Expression Recognition Based on Fusion of Gabor and LBP Features", *Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition*, Hong Kong, 2008

[6] 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.

[7] Hanchuan Peng, Fuhui Long, and Chris Ding, "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.

[8] Espinosa-Duro V., Faundez-Zanuy M., "Face Identification by Means of a Neural Net Classifier," *Proceedings of IEEE 33rd Annual 1999 International Carnahan Conf. on Security Technology*, pp. 182-186,1999.

[9] Dasarathy B.V., "Nearest Neighbor Norms: NN Pattern Classification Techniques.",*IEEE Computer Society Press,* Los Alamitos, CA, 1991.

[10] Smith, L.I., "Tutorial on Principal Component Analysis," 2002.

[11] Lajevardi S.M., Lech M., "Facial Expression Recognition Using a Bank of Neural Networks and logarithmic Gabor Filters", *DICTA08*, Canberra, Australia, 2008.

[12] Lajevardi S.M., Lech M., "Averaged Gabor Filter Features for Facial Expression Recognition", *DICTA08*, Canberra, Australia, 2008.

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