ANEXPOSITION ANALYSIS OF FACIAL EXPRESSION RECOGNITION TECHNIQUES

A. H. Butt, M. K. Mahmood and Y D Khan

Department of Computer Science, School of Science and Technology, University of Management and Technology, Lahore
Department of Mathematics, University of the Punjab, Lahore
Corresponding author e-mail: ahmad.hassan@umt.edu.pk

ABSTRACT: This study compared the technique such as Gabor-wavelet filter banks and Local Binary Patterns (LBP) that were commonly used for facial expression recognition. Broad experimental phases were designed to analyze the accuracy and efficiency of each state of art techniques regarding facial expression recognition. A close analysis of the Receiver Operating Characteristic (ROC) and overall accuracy of the results for well-known facial databases revealed that the LBP-based extraction of facial expression features was the most efficient and effective technique.

Key words: Facial expression recognition, Gabor filters, Local binary patterns, support vector machines.

(Received 15-09-2015 Accepted 10-09-2016)

INTRODUCTION

A multitude of information can be derived from the face of human beings during any sort of interaction. In casual and social interaction the most usualtransmitting source is the facial expression of the humans. Computer-based facial expression extraction for identification and prediction analysis has been a dynamic topic in the community of the research over two decades (Valstar et al., 2005). Facial expression provides information about one’s personality, emotions and thoughts (Belhumeur et al., 1997). Facial expressions carry information regarding the human thoughts and temperaments which might not always be expressed through words. Furthermore, it plays an important role in human communication and nonverbal interactions as well as delicate signs about an enthusiastic reaction (Lien and James, 1998).

Facial expressions supplement verbal correspondence, or can pass on complete contemplations independent from any other form of expression. Subsequently, facial expression analyses has drawn substantial attention in the progress of human computer interaction as it acts as a natural and effective source of communication among humans (Zhang et al., 1998). The advancements in pattern recognition techniques have helped scientists to develop intelligent algorithms to facilitate and improve human-computer interactive interfaces (Liu et al., 2009).

Individuals possess a unique facial expression of distinctive content altogether with six basic primary emotions (Ekman and Friesen, 1977). These prototypes are referred as the basic emotional displays invariant over human ethnicity and societies. These emotional displays are generally enumerated as anger, sadness, disgust, fear, surprise and happiness (Fasel and Luettin, 2003). Although great results have been achieved but determining facial expressions with a high accuracy stays difficult due to the nuance, intricacy and variety of facial variances.

A wide range of applications are based on determination of human facial expression such as surveillance, user identification, psychological analysis and facial animations. Certain intelligent applications have been developed which observe a conversation and use the facial expressions of the subject to gauge the responses of the user regarding like or dislike of the specific components of the conversation (Pantic and Rothkrantz, 2000). Determining efficiently facial expression from facial images is a crucial task for effective face expression prediction (Shan et al., 2009). Two common approaches were mainly used to find facial expressions: geometric points based techniques and key appearance based approach. Geometric features made use of the location and the shape of the components of the facial image which after extraction were populated into a feature vector representing the face geometry (Shan et al., 2009).

In a study carried out by Gabor wavelets used as feature extraction technique based on the principle of image masking. It uses either full or partial facial image to find the feature variations considering the enhanced performance of Gabor filters. The majority of the scientists who have worked on feature-based techniques have preferred the use of Gabor-wavelet filters (Shan et al., 2009). Facial Expression conveys non-verbal communication cues in face-to-face interactions. A Facial Action Coding System (FACS) was designed for describing visually distinguishable facial movements (Ekman and Friesen, 1978). Using the FACS, action parameters were designated to each of the Expression which classify the human emotions. Also it was demonstrated that the verbal portion of a message has contributions up to 7 percent in the total effect of the original message, the facial expression part adds up to
55 and vocal part contributes 38 percent (Mehrabian, 1968).

Several researchers have presented many solutions to the problem using different mathematical or statistical models (Khan et al, 2012 and Khan et al, 2014). Darwin worked on different expressions of humans relating to various emotions (Ekman, 2006). Since then, facial expression recognition has been the area of research for many scientists. Some scientist put forward the idea of automatically analyzing facial expressions from images (Suwa et al., 1978). There has been a lot of advancement in the last few years to find an improved mechanism which evaluates facial expressions. A detailed study of the existing work can be found in the work conducted by (Fasel and Luetten, 2003, Cohen et al., 2003 and Pantic and Rothkrantz, 2000). In the current context a review is presented for some of the previous work proposed by various researchers.

Face expression recognition is an important and tough area as several encumbrances and challenges are encountered. The most basic difficulty arises in identifying the neutral expression as they are generally the starting point. The neutral expression typically looks too rudimentary and the changes between them are quite difficult to find. Thus, facial images of the front view form a very large and congested cluster in image space. Traditional pattern recognition techniques are rendered incapacitated to discriminate among intermediate expression in order to identify the starting point with a high degree of success (Ryan et al., 2009). Furthermore, the nature of human face is not identifiably rigid. Furthermore, there are many aspects that form the appearance of the face. The sources of the images in the facial expression cannot be categorized into groups if the quality of the image is appropriate (Emerich et al., 2009).

MATERIALS AND METHODS

The importance of facial features for facial expression recognition cannot be overstated. Many face expression recognition systems required facial expression features in addition to the holistic face, as suggested in a study in psychology (Ahonen et al., 2006). Images from various databases were used for experimental purposes. Firstly, some of the popular databases were illustrated. There were two databases which were commonly used by researchers: the Cohn-Kanade (CK) AU-Coded face expression image database (Tian et al., 2001) and Japanese Female Facial Expression Database (JAFFE) database.

Cohn-Kanade AU-Coded Face Expression Image Database— it has been used in most of the areas where automatic facial expression analysis and recognition was required. It contained images of over 200 university students between ages 18 to 30 years, 70 percent were females and 30 percent were males, 18 percent were African-American, and 10 percent were Asian-Americans. Different combinations of facial expressions from these subjects were collected to compile a database. Images were taken in a special observation room with stable lighting and consistent backgrounds (Tian et al., 2001).

A total of 540 image sequences were chosen from the database which started with a neutral expression. These images were converted into 640 x 480 pixel resolution with 8-bit grayscale encoding. The image sequences of 90 subjects were recorded with all the basic six emotions per subject which included emotion expressions like joy, surprise, fear, anger, sadness, disgust while some were neutral (Tian et al., 2001).

The Japanese Female Facial Expression Database-- This database included 213 images of 10 Female Japanese models along with 7 different facial expressions. Six of the expressions were of basic nature i.e. anger, disgust, fear, sad, happy, and surprise and the last one was neutral. The database was provided by (Lyonset et al., 1998). The images were collected by the Psychology Department at Kyushu University. The resolution of the images was 256 x 256 pixels.

Fig. 1. The sample images of face with different expressions from the Cohn-Kanade database
In a study conducted by (Zhang et al., 1998), two types of feature extraction operations were performed. The first type of feature extraction used fiducial points based on geometric positions and the second type was based on multi-scale Gabor wavelet coefficients which were extracted, using fiducial points from the facial image. Each image was represented by 34 fiducial points which were manually chosen. These 34 coordinate points are shown in Fig. 3 which were used as features in their image analysis.

**Gabor filters:** Gabor filters, named after Dennis Gabor, were mostly used in the field of image processing. Gabor filters were based in terms of parameters like frequency and orientation and hence had parallels with the human visual perception system.

The sinusoidal plane wave modulation of Gaussian kernel function in the spatial domain was constructed as a 2D Gabor filter. The Gaussian function multiplied by a sinusoidal wave acted as a plane wave in case of 2D Gabor filters was measured as its impulsive response. The Gabor filter represented orthogonal focusing points as a component of two parts which were real and imaginary. A complex number was formed from these two components. The following equations of Gabor filters in 2D plane were formed by appropriate dilation and rotation in the mother Gabor wavelet function (Kobayashi and Hara, 1992).

\[
g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x^2 + \gamma^2 y^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right)
\]

**Equation 1: 2D Gabor Filter function for the real component**

\[
g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x^2 + \gamma^2 y^2}{2\sigma^2}\right) \sin\left(2\pi \frac{x'}{\lambda} + \psi\right)
\]

**Equation 2: 2D Gabor Filter function for the imaginary component**

![Fig. 2. The sample images of Japanese models with expressions from the JAFFE database](image)

![Fig. 3. The 34 fiducial points of facial geometry](image)

![Fig. 4. The Gabor filters with different lengths, sizes and measurable-points.](image)
g(x, y; λ, θ, ψ, σ, γ) = \exp \left( -\frac{x'^2 + y'^2}{2\sigma^2} \right) \exp \left( i \left( 2\pi \frac{x'}{\lambda} + \psi \right) \right)

**Equation 3: 2D Gabor Filter combined function as the complex component**

Where

\[ x' = x \cos \theta + y \sin \theta, \quad y' = -x \sin \theta + y \cos \theta \]

In the above equations, the sinusoidal factor wavelength was represented as \( \lambda \) whereas the orientation of the normal to the parallel stripes of a Gabor function was represented by \( \theta \), the phase offset was represented by \( \psi \), \( \sigma \) was the standard deviation and the spatial aspect ratio is represented as \( \gamma \), through which the ellipticity of the support of the Gabor function is specified.

In a study conducted by (Bashyalet et al., 2008) Gabor filter banks were successfully used to extract features from the facial images. The results of the study proved that the accuracy of the system was no better than 90.22\% in correctly classifying facial expressions.

This technique did not work well with the erratic nature of real life data which was highly involved in applications used by high level security centers for quality surveillance which was based on facial expression recognitions.

**Binary pattern:** Originally the LBP operator was first used by (Ojala et al., 1996), and it came out as a strong choice for texture description. The operator used a 3 x 3 pixel image as a threshold in the neighborhood of every pixel in the image it replaced the kernel value comparison with the neighboring values. This resulted as a binary number and a histogram of the LBP labels which were calculated over the region of the images used as a texture descriptor. The histogram of the LBP image labels based on 256-bin were later used for template matching identifications.

![Fig. 5. The image of the face, which is divided into 3 x 3 regions](image)

**Fig. 6. The LBP operator used to create a template image.**

The LBP operator had proven to be supportive and its major benefits were its ability to detect change in gray level intensity and deficiency in computation. These features ensured preference for image analysis problems. The LBP operator has been extended making it capable to use template of different texture patterns at different scales and different sizes of neighborhoods (Ojala et al., 2002). Neighborhood template image was characterized as a set of points based on LBP operator. The LBP operator labeled images were used in the histogram transformation. The labeled image \( f(x, y) \) was defined as Histogram in the following equation,

\[
H = \sum_{x,y} \mathbb{I}(f(x,y) = t), \quad t = 0, \ldots, l - 1
\]

Where \( \mathbb{I} \) represented the variable labels generated by the LBP operator.

**Support vector machines** Machine learning algorithms encountered in data inputs throughout a training stage. These algorithms created a model of the hypothesis function of the input and output to predict future data. A template-based approach was used in previous work to detect the facial regions in images (Ahonen et al., 2004). In this method a template image was formed that contained face regions. The nearest-neighbor-classifier was used for the matching of the template with the nearest testing image. Another successful approach to recognize facial expression image was the use of Support Vector Machines as has been reported by (Cristianini and Shawe, 2000 and Murthy and Jordon, 2009) based on the principle of statistical learning. SVMs embed the data into a feature space of very high dimension, in which geometry and linear algebra were used to disseminate data. Table 1 showed the experimental results of applying SVM on the LBP operated images. About 200 test images from JAFFE database which were used to test the system.

**Neural Networks:** In the recent years, some researchers employed the use of neural networks for classifying facial expressions. The work of (Kobayashi and Hara, 1992 and
Kobayashi and Hara, 1994) proposed neural networks based models but the techniques used to extract facial features were not very effective and required manual intervention. To address such shortcomings, (Chang and Chen, 1999) proposed a fully automated expression recognition system built through an accurate facial feature extraction technique. In this technique radial basis function network (RBFS) and multilayer perceptron (MLP) neural network models were used. The authors used six efficient Action Units (AUs) as inputs to the neural classifier as compared to the usage of 60 inputs of x and y coordinates formed by 30 facial fiducial points. The experimental results obtained from the tests produced recognition rates as high as 92.1% which was quite adequate recognition rate. A neural network emulated the neural cognitive system in humans. A multilayer neural network typically had an input layer, a hidden layer and an output layer as is shown in figure 8.

Table 1 LBP based Facial expression extraction confusion matrix using 7-class SVM classifier.

<table>
<thead>
<tr>
<th>Actual Features</th>
<th>Joy (%)</th>
<th>Surprise (%)</th>
<th>Fear (%)</th>
<th>Anger (%)</th>
<th>Sad (%)</th>
<th>Disgust (%)</th>
<th>Neutral (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Surprise</td>
<td>3.33</td>
<td>96.67</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
<td>3.03</td>
<td>90.31</td>
<td>0</td>
<td>0</td>
<td>3.33</td>
<td>3.33</td>
</tr>
<tr>
<td>Anger</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sad</td>
<td>0</td>
<td>0</td>
<td>3.33</td>
<td>0</td>
<td>93.34</td>
<td>0</td>
<td>3.33</td>
</tr>
<tr>
<td>Disgust</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.33</td>
<td>0</td>
<td>93.34</td>
<td>3.33</td>
</tr>
<tr>
<td>Neutral</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

Overall Average: 96.23%

Figure 7: Two classes are represented by disks and circles in $\mathbb{R}^d$. The OSH (Optimal Separating Hyper-Plane) is constructed by SVM which determined the margins between two classes maximum by the input space $\mathbb{R}^d$ mapping onto a higher dimension which is the feature space $H$. The kernel function $K(\tilde{x}_i, \tilde{x}_j)$ is used to determine the mapping.

Figure 8: A Multilayer Neural Network
RESULTS AND DISCUSSION

Different experiments were conducted on each database to test the robustness, accuracy and efficiency of the previously described models. The images were convolved through a Gabor Filter. The generated feature vector was further analyzed to classify the facial expressions. In subsequent experiments same database was used. LBP operator was applied on each using a 3x3 pixel window as threshold. The LBP operator replaced the center value in comparison with the neighboring pixels. Ultimately the formed histogram was used as a feature vector to identify facial expressions embedded in each image. Further, SVM was used as a classifier. Feature vectors extracted from each image database were fed into an array of multiple SVMs. Each SVM was trained for different expressions. Subsequently, neural networks were also used. A large subset of all the feature vectors was used for training of neural network while some samples were adequately left for testing. Later, the trained neural network was used for classification of facial expressions. The results from these experiments were collected and tabulated into a confusion matrix which depicted the overall accuracy of each system. Table 1-3 summarized the results showing the percentage of positive matches diagonally.

Table 2. Gabor Filter based Facial expression extraction confusion matrix.

<table>
<thead>
<tr>
<th>Actual Features</th>
<th>Predicted Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy ( % )</td>
<td>Surprise ( % )</td>
</tr>
<tr>
<td>Joy</td>
<td>89.56</td>
</tr>
<tr>
<td>Surprise</td>
<td>0</td>
</tr>
<tr>
<td>Fear</td>
<td>5.45</td>
</tr>
<tr>
<td>Anger</td>
<td>0</td>
</tr>
<tr>
<td>Sad</td>
<td>0</td>
</tr>
<tr>
<td>Disgust</td>
<td>5.76</td>
</tr>
<tr>
<td>Neutral</td>
<td>0</td>
</tr>
</tbody>
</table>

Overall Average: 90.22%

Table 3. Neural Networks based Facial expression extraction confusion matrix.

<table>
<thead>
<tr>
<th>Actual Features</th>
<th>Predicted Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy ( % )</td>
<td>Surprise ( % )</td>
</tr>
<tr>
<td>Joy</td>
<td>92.61</td>
</tr>
<tr>
<td>Surprise</td>
<td>2.25</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
</tr>
<tr>
<td>Anger</td>
<td>0</td>
</tr>
<tr>
<td>Sad</td>
<td>3.14</td>
</tr>
<tr>
<td>Disgust</td>
<td>2.81</td>
</tr>
<tr>
<td>Neutral</td>
<td>2.87</td>
</tr>
</tbody>
</table>

Overall Average: 92.1%

Moreover, Table-4 shows the overall accuracy yielded from each of the described model. The overall accuracy of each system is depicted in figure 9 as a histogram.

Table 4 Comparison of different methods used in Facial Expression Recognition System.

<table>
<thead>
<tr>
<th>Methods/Techniques</th>
<th>Results/Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Network Classifiers</td>
<td>92.1%</td>
</tr>
<tr>
<td>LBP based SVM</td>
<td>96.23%</td>
</tr>
<tr>
<td>Gabor Filters</td>
<td>90.22%</td>
</tr>
<tr>
<td>PCA with Eigen Faces</td>
<td>83%</td>
</tr>
</tbody>
</table>

Figure 9: Histogram of the different techniques with measures of accuracy
Facial images of 7 different facial expression were collected from Cohn-Kanade (CK) AU-coded face expression image database (Tian et al., 2001) and Japanese Female Facial Expression Database (JAFFEE) database (Lyons et al., 1998). Some of the most assiduous features extraction and classification models were portrayed in this work. Bank of Gabor filters and LBP described by (Kobayashi and Hara, 1992 and Ojala et al., 2002) respectively were used for feature extraction. The classification models outlined were based on SVM, Gabor filters and Neural Networks as described by (Cristianini and Shawe, 2000, Murthy and Jadon, 2009 and Bashyal et al., 2008) respectively.

Usually data used for training a model was obtained from the outcome of previous experiments. Sometimes depending upon a phenomenon it might not be possible to readily obtain new experimentally proven data for testing the accuracy of prediction model. Also, some authentic data might be available but it might be scarce and insufficient to reach a conclusive result. In such a situation certain exhaustive tests were performed to determine the credibility of the model. In this section two such tests i.e. cross-validation and jackknife tests were described. Cross-validation was a way to expect that the purposed method was perfect when an obvious validation set was not available as described in the work conducted by (Fang et al., 2008). The available data was split into k-folds where k was some constant. All the partitions are disjoint. The system was tested for each partition while it was trained for the rest of data. The test was iterated for k=10 for every random disjoint partition. The overall average of the accuracy in each iteration was reported as the cross-validation result. Jackknife testing is one of the most commonly used and mature resampling technique also described by (Feng et. al., 2008). Other validating techniques used randomly selected or partitioned dataset for testing predictor. Usually there were no rules that could govern the partitioning of this data. The data could be partitioned in many different ways hence it was possible that a certain partition may produce good results while another partition may not behave likewise. In such subsampling technique very small selection was used for testing. Consequently, different selection might give entirely different results. Therefore, such methods might never produce unique results. The strength of jackknifing lies in its ability to produce unique results. The jackknife method computed the overall accuracy of the predictor by thoroughly leaving out each observation from a dataset and calculating the accuracy using left out data. Table-5 describes the result of 10-fold cross validation and jackknife testing performed on each of the described techniques.

### Table 4. Validation testing results for each of the described technique.

<table>
<thead>
<tr>
<th>Methods</th>
<th>10-Fold Cross Validation</th>
<th>Jackknife testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural network classifiers</td>
<td>91.0%</td>
<td>90.33%</td>
</tr>
<tr>
<td>LBP based SVM</td>
<td>95.01%</td>
<td>94.89%</td>
</tr>
<tr>
<td>Gabor filters</td>
<td>89.3%</td>
<td>88.73%</td>
</tr>
<tr>
<td>PCA with Eigen faces</td>
<td>82.3%</td>
<td>81.9%</td>
</tr>
</tbody>
</table>

Automatically extracting and predicting facial expressions werea vital task in order to describe human emotions and para-linguistic communication. Such applications also helped to create multi-modal graphical interfaces and applications that were related to security and surveillance. This study presented different...
methodologies which were used in the past to enhance the performance of various facial expression recognition techniques. It was compared with LBP methods with Gabor-wavelet technique and have concluded that LBP was better approach in feature extraction methods used for facial expressions. Results clearly demonstrate that images of facial expressions were considered as a collection of micro template such as lines and edges which were well defined by Local Binary Patterns. PCA based techniques could only achieve 83% as the best possible recognition result. Using LBP methods in facial expression techniques had produced better performance results of 96.23% on average, which was better than 90.22% of the Gabor-filter based technique used by (Bashyal and Venayagamoorthy, 2008). Neural network based classifiers worked no better than 92.1% in (Chang and Chen, 1999). This study mainly contributed to investigate the various facial expression recognition techniques based on the static image analysis and video based image features. The results shown in table 1-4 suggested that LBP-SVM is the most accurate method for facial expression recognition.

REFERENCES


