CHAPTER 8

OPTICAL NEURAL NETWORK BASED FACE EXPRESSION RECOGNITION

8.1 INTRODUCTION

Humans detect and interpret faces and facial expressions in a scene with little or no effort. There are several related problems: detection of an image segment as a face, extraction of the facial expression information, and classification of the expression (e.g., in emotion categories). A system that performs these operations accurately and in real time would form a big step in achieving a human-like interaction between man and machine.

Relatively few existing works combine different modalities into a single system for human communicative reaction analysis. Examples are the works of Chen (1997) and Donato (1999) who studied the effects of a combined detection of facial and vocal expressions of emotions.

Recent advances in image analysis and pattern recognition open up the possibility of automatic detection and classification of emotional and conversational facial signals. Automatic facial expression analysis could bring facial expressions into man-machine interaction as a new modality and make the interaction tighter and more efficient. Such a system could also make classification of facial expressions widely accessible as a tool for research in behavioral science and medicine.
Kimura and Yachida (1997) fit a Potential Net to each frame of the examined facial image sequence. The pattern of the deformed net is compared to the pattern extracted from an expressionless face (the first frame of a sequence) and the variation in the position of the net nodes is used for further processing.

In the proposed approach first step is to detect the image as a face or a non-face using optical neural network. After detecting a face it is built as an emotion space by applying KLT on the images, of expressions like surprise, sadness, fear, angry, smiling, and disgust shown by a single person gradually, from expressionless to a maximum. The eigen space spanned by the first three principal components has been used as the emotion space, onto which an input image is projected for a quantified emotional classification.

To explore the issues in design and implementation of a system that could perform automated facial expression analysis. In general, three main steps can be distinguished in tackling the problem. First, before a facial expression can be analyzed, the face must be detected in a scene. Next is to devise mechanisms for extracting the facial expression information from the observed facial image or image sequence. In the case of static images, the process of extracting the facial expression information is referred to as localizing the face and its features in the scene.

After the presence of a face has been detected in the observed scene, the next step is to extract the information about the encountered facial expression in an automatic way. If the extraction cannot be performed automatically, a fully automatic facial expression analyzer cannot be developed. Both, the applied face representation and the kind of input images affect the choice of the approach to facial expression information extraction.
After the face and its appearance have been perceived, the next step of an automated expression analyzer is to identify the facial expression conveyed by the face. A fundamental issue about the facial expression classification is to define a set of categories we want to deal with. A related issue is to devise mechanisms of categorization. Facial expressions can be classified in various ways like surprise, sadness, fear, angry, smiling, and disgust.

Fully automatic and real-time facial expression recognition could find many applications, for instance, in human-computer interaction, biometrics, telecommunications, and psychological research. Most of the research on facial expression recognition has been based on static images. We propose a novel, theoretically and computationally simple approach based on Gabor features and detection using optical neural networks. The features extracted are not only insensitive with respect to translation and rotation, but also robust with respect to monotonic gray-scale changes caused, for example, by illumination variations. This will make our approach a highly valuable tool for many potential computer vision applications.

### 8.1.1 Proposed approach

The proposed approach for face recognition and its expression using optical neural networks is detailed in the block diagram shown in Figure 8.1.

![Figure 8.1 Block Diagram for Face Recognition and its expression](image-url)
8.1.2 Feature Extraction

The intensity of an image is the only source from a camera used for object recognition. However, a lot of variations, such as color and shape of the object, lighting, etc., are all encoded as the intensity. To eliminate the extrinsic factors, various feature extraction and selection methods are widely used. For almost three decades the use of features based on Gabor filters has been promoted for their useful properties in image processing. The most important properties are related to invariance to illumination, rotation, scale, and translation. These properties are based on the fact that they are all parameters of Gabor filters themselves. This is especially useful in feature extraction, where Gabor filters have succeeded in many applications, from texture analysis to iris and face recognition.

The Gabor wavelet was first introduced by David Gabor in 1946. The Gabor wavelet is a sinusoidal plane wave with a particular frequency and orientation, modulated by a Gaussian envelope. It can characterize the spatial frequency structure in the image while preserving information of spatial relations and, thus, is suitable for extracting the orientation-dependent frequency contents of patterns.

Also, the use of Gabor filters in extracting textured image features is motivated by various factors. The Gabor representation has been shown to be optimal in the sense of minimizing the joint two-dimensional uncertainty in space and frequency. These filters can be considered as orientation and scale tunable edge and line (bar) detectors, and the statistics of these micro features in a given region are often used to characterize the underlying texture information. Gabor features have been used in several image analysis applications including texture classification and segmentation, image recognition, image registration, and motion tracking.
Lades (1993) pioneered the use of the Gabor wavelet for face recognition using the dynamic link architecture framework. Wiskott (1997) subsequently developed a Gabor wavelet-based elastic bunch graph matching (EBGM) method to label and recognize human faces. In the EBGM method, the face is represented as a graph, each node of which contains a group of coefficients, known as a jet. It can also measure the geometry of the face by using the labeled distance vector, which is the edge part of the graph. Liu and Wechsler (2002) showed that the face representation based on the magnitude part of Gabor feature had been a promising way towards achieving high accuracy face recognition. Zhang (2005) provided an AdaBoost-based strategy to select the discriminative features from the magnitude part of the Gabor feature, and then trained a Fisher classifier to make a final classification. As a powerful descriptor, the Gabor wavelet is also used in many applications, such as data compression optical character recognition (OCR), texture analysis, fingerprint recognition, and so on. Most of the above applications are based on the magnitude part of Gabor feature. In fact, the Gabor phase is a very discriminative information source, and has been successfully used in iris and palm print identification.

8.1.3 Gabor Functions and Wavelets

A two dimensional Gabor function \( g(x, y) \) and its Fourier transform \( G(u, v) \) can be written as:

\[
\begin{align*}
g(x, y) &= \left(\frac{1}{2\pi \sigma_x \sigma_y}\right) \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\sigma W_x\right] \\
G(u, v) &= \exp \left\{-\frac{1}{2} \left[\frac{(u - W_u)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]\right\}
\end{align*}
\]  

(8.1)  

(8.2)
where $\sigma_{x} = \frac{1}{2\pi} \sigma_{x}$ and $\sigma_{y} = \frac{1}{2\pi} \sigma_{y}$. Gabor functions form a complete but nonorthogonal basis set. Expanding a signal using this basis provides a localized frequency description. A class of self-similar functions, referred to as Gabor wavelets in the following discussion, is now considered. Let $g(x, y)$ be the mother Gabor wavelet, then this self-similar filter dictionary can be obtained by appropriate dilations and rotations of $g(x, y)$ through the generating function:

$$g_{mn}(x, y) = a^{-m}G(x', y'), a > 1, \ m, n = \text{integer} \quad (8.3)$$

$$x' = a^{-m} (x \cos \theta + y \sin \theta), \quad \text{and} \quad (8.4)$$

$$y' = a^{-m} (-x \sin \theta + y \cos \theta), \quad (8.5)$$

where $\theta = n\pi / K$ and $K$ is the total number of orientations. The scale factor $a^{-m}$ is meant to ensure that the energy is independent of $m$.

### 8.1.4 Optical Neural Network

The schematic diagram of a compact size LCTV optical neural network is shown in Figure 8.2. The ONN is designed with $8 \times 8 = 64$ fully interconnected neurons in which a Xenon arc Lamp was used as the incoherent light source. LCTV 1 is for displaying the IWM, which consists of an $8 \times 8$ array of sub matrices, and each sub matrix has $8 \times 8$ elements. This IWM is then displayed on a fine diffuser immediately behind LCTV1. The lens let array, which consists of $8 \times 8$ lenses, provides the interconnections between the IWM and the input pattern. Each lens of the lens let array images each of the IWM sub matrices onto the input LCTV2 to establish the interconnections. Thus the input matrix (i.e., LCTV2) is interconnected with all the sub matrices. The output result from LCTV2 is then collected by a
lens, which images the lenslet array onto the CCD camera. The signals detected by the CCD are sent to a threshold circuit and the final results can be fed back to the LCTV2 for next iteration. Thus the proposed LCTV optical neural network is an adaptive neural network.

**Figure 8.2 Optical neural networks**

Each Neuron in the network is composed of an input summing port, a nonlinear transfer device, and an output port. In an opto-electronic system, a differential pair of detectors is operated as the input to the neuron; signals with positive (excitatory) weights arrive at one detector, and signals with negative (inhibitory) weights arrive at the other detector. These detectors sum up the intensity at each optical signal arriving at the neuron. The neuron’s activation function is electronically applied to the detected signal to produce an output signal. The output signal drives either an optical source (or pair of sources) or a spatial light modulator. Figure 8.3 illustrates an opto-
electronic neuron that uses pair of laser diodes to encode the neuron’s output signal. Figure 8.4 illustrates the feed-forward optical neural network system.

Figure 8.3 Opto-electronic neuron constructed from detectors and emitters

Figure 8.4 Opto-electronic feed forward neural network system
The steps followed for training the optical neural network are outlined as follows:

1. Create an initial set of non-face images by generating 1000 random images.
2. Train the optical neural network using the extracted gabor features to produce an output of +1 for the face examples, and -1 for the nonface examples. In the first iteration, the network’s weights are initialized random. After the first iteration, we use the weights computed by training in the previous iteration as the starting point.
3. Run the system on an image of scenery which contains no faces.

Collect sub images in which the network incorrectly identifies a face.

![Figure 8.5 Different facial expressions in Cohn database.](image)

The architecture for the identifier network consists of three layers, an input layer of 2160 units, a hidden layer of 90 units, and an output layer of single unit. Each unit uses a hyperbolic tangent activation function, and the network is trained using the scaled conjugate gradient back propagation algorithm. Testing patterns having 800 images are done to validate the proposed method.

8.1.5 KLT for facial expression recognition

Karhunen–Loeve Transform based dimensionality reduction for face images was first proposed by Kirby and Sirovich (1990). Mathematically,
the eigenface method tries to represent a face image as a linear combination of orthonormal vectors, called eigenfaces. These eigenfaces are obtained by finding the eigenvectors of the covariance matrix of the training face image set. Let \( I_1, I_2, I_3, \ldots, I_k \) be a set of \( k \) face images, each ordered lexicographically.

The eigenvectors of the matrix

\[
C = \sum_{i=1}^{K} I_i I_i^T
\]  

(8.6)

that corresponds to the largest eigenvalues span a linear subspace that can reconstruct the face images with minimum reconstruction error in the least squares sense. This \( L \)-dimensional subspace is called the face space.

Assuming is a lexicographically ordered face image and is the matrix that contains the eigenfaces as its columns, we can write

\[
x = \phi a + e_x
\]  

(8.7)

where ‘\( a \)’ is the feature vector that represents the face, and \( e_x \) is the subspace representation error for the face image. As a larger training data set is used and the dimensionality of the face space is increased, the representation error gets smaller. Letting

\[
a = [a_1, a_2, \ldots, a_L]^T
\]  

(8.8)

be the feature vector, and

\[
\phi = [\phi_1, \phi_2, \ldots, \phi_L]
\]  

(8.9)
be the matrix where are the eigenface vectors, is computed as follows:

\[ a_i = \phi_i^T x . \]  

(8.10)

The computed eigen values are compared for similarity and the expression of the face is recognized.

8.2 EXPERIMENTS

The aim of these experiments is to study the robustness, of the proposed facial expression recognition approach, to the occlusion of face regions. The experiments have been carried out using image sequences from the Cohn-Kanade (2000) Action-Unit-coded Carnegie Mellon University facial expression database, where many sequences are available for each expression class. The Cohn-Kanade facial expression database consists of 100 university students ranging in age from 18 to 30 years. Sixty-five percent were female, 15 percent were African-American, and 3 percent were Asian or Latino. The subjects were instructed by an experimenter to perform a series of 23 facial displays that included a single Action Unit and combinations of Action Units, six of which were based on descriptions of prototypical emotions of anger, disgust, fear, joy, sadness, and surprise. The example set of various face expression patterns are shown in Figure 8.5. The image sequences from neutral to the target display were digitized into 120 x 100 pixel arrays with an 8-bit precision for gray-scale values.

<table>
<thead>
<tr>
<th>Table 8.1 Table for classifying the emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sadness</td>
</tr>
<tr>
<td>0 &lt; ( \lambda ) &lt; 10</td>
</tr>
</tbody>
</table>

In Table 8.1 \( \lambda \) represents eigen value.
Figure 8.6 Classification Percentage of various expressions

Table 8.2 Comparison of the proposed approach with the existing systems

<table>
<thead>
<tr>
<th></th>
<th>Surprise</th>
<th>Happiness</th>
<th>Sadness</th>
<th>Fear</th>
<th>Anger</th>
<th>Disgust</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP-TOP(_{4,4,4,3,3,3})</td>
<td>98.65</td>
<td>89.11</td>
<td>90.41</td>
<td>83.93</td>
<td>93.75</td>
<td>92.11</td>
<td>91.18</td>
</tr>
<tr>
<td>LBP-TOP(_{8,8,8,3,3,3})</td>
<td>100.00</td>
<td>97.03</td>
<td>94.52</td>
<td>85.71</td>
<td>84.38</td>
<td>97.37</td>
<td>94.38</td>
</tr>
<tr>
<td>LBP-TOP(_{8,8,8,1,1,1})</td>
<td>97.30</td>
<td>90.10</td>
<td>89.04</td>
<td>82.14</td>
<td>84.38</td>
<td>97.37</td>
<td>90.37</td>
</tr>
<tr>
<td>LBP-TOP(_{16,16,16,3,3,3})</td>
<td>98.65</td>
<td>91.10</td>
<td>89.04</td>
<td>76.80</td>
<td>75.00</td>
<td>92.10</td>
<td>88.77</td>
</tr>
<tr>
<td>VLBP(_{3,2,3})</td>
<td>95.95</td>
<td>94.06</td>
<td>89.04</td>
<td>83.93</td>
<td>87.50</td>
<td>92.10</td>
<td>91.18</td>
</tr>
<tr>
<td>LBP-TOP+VLBP</td>
<td>100.00</td>
<td>97.03</td>
<td>94.52</td>
<td>89.29</td>
<td>87.50</td>
<td>97.37</td>
<td>95.19</td>
</tr>
<tr>
<td>Proposed System</td>
<td>98.1</td>
<td>95.2</td>
<td>95.8</td>
<td>92.7</td>
<td>94.2</td>
<td>95.5</td>
<td>95.24</td>
</tr>
</tbody>
</table>

The optical neural net work is trained with 1000 patterns and tested for 800 patterns and validated. The approach is quite robust with respect to variations of illumination. The experimental results of the proposed approach are given in Figure 8.6 and Table 8.2 gives a concise presentation comparing
with other approach analyzed by Guoying (2007). The result shows that our approach outperforms the other dynamic and static methods.

In the proposed approach, a face expression recognition system using optical neural network and KLT is presented. Novel algorithm optical neural network for face detection and localization is proposed, that exploit depth information to achieve robustness under background clutter, occlusions, face pose, and illumination variations. Face expression recognition is based on the well-known KLT method, which is also applied to identify the expressions based on eigen pattern. Unlike most methods in face recognition literature that perform recognition or expression the proposed method simplifies the processing of image. From an inspection of the experimental results presented in Table 8.1, it can be concluded that the verification performance of the approach is better.

8.3 OVERALL EXPERIMENTAL ANALYSIS

A neural network as well as optical neural network architectures is designed for rotation invariant face recognition and facial expression recognition.

In the first experiment, face recognition is done using neural networks. Initially, human faces are identified by using identifier network and Karhuenen-Loeve Transform is used for recognition.

The system does well on the upright test set. A classification rate of 91.3% is obtained. But it has a poor detection rate on the rotated test set. A classification rate of 17% is obtained.
To check the capability of the proposed approach, the rotated faces are made upright by using orientation network and then applied to the identifier network. The recognition is performed by using Karhuenen-Loeve Transform. This system does well on both upright test and rotated test set. A classification rate of 91.8% for upright test set and 88.5% for the rotated test is obtained. It is proved that, in the proposed system, the classification results are improved.

In the second experiment, the optical neural network based orientation network and identifier network are implemented. The above experiments are repeated with the same real database. A classification rate of 92.8% for the upright test and 90.4% for the rotated test set is obtained. From the results of our second experiment, it is proved that, the recognition rate is improved by using optical neural networks.

Figure 8.7  Comparison between face recognition using neural network and optical neural network for upright test
Thirdly, the experiments are analyzed for facial expression recognition from Cohn Kanade database by using optical neural network and Karhuenen-Loeve Transform. The optical neural network is trained with 1000 patterns and tested for 800 patterns and validated. The six facial expressions recognized are surprise, happiness, sadness, fear, anger and disgust.

The results are compared with other approaches, analyzed by Guoying (2007). The result shows the proposed approach which outperforms the other dynamic and static methods.

8.4 SUMMARY

The facial expression recognition using optical neural network is presented. The six basic expressions, surprise, happiness, sadness, fear, anger and disgust are taken and experiment is conducted. The results are tabulated and it is compared with previous existing techniques. The result shows that, the proposed system outperforms the existing techniques.