

# Automatic Recognition of Smiling and Neutral Facial Expressions

P. Li, S. L. Phung, A. Bouzerdom, and F. H. C. Tivive  
School of Electrical, Computer and Telecommunication Engineering,  
University of Wollongong, Wollongong, NSW 2522, Australia

pli, phung, bouzer, tivive@uow.edu.au

## Abstract

*Facial expression is one way humans convey their emotional states. Accurate recognition of facial expressions via image analysis plays a vital role in perceptual human-computer interaction, robotics and online games. This paper focuses on recognising the smiling from the neutral facial expression. We propose a face alignment method to address the localisation error in existing face detection methods. In this paper, smiling and neutral facial expression are differentiated using a novel neural architecture that combines fixed and adaptive non-linear 2-D filters. The fixed filters are used to extract primitive features, whereas the adaptive filters are trained to extract more complex features for facial expression classification. The proposed approach is evaluated on the JAFFE database and it correctly aligns and crops all images, which is better than several existing methods evaluated on the same database. Our system achieves a classification rate of 99.0% for smiling versus neutral expressions.*

## 1. Introduction

Facial expression, controlled by a complex mesh of nerves and muscles beneath the face skin, enables people to convey emotions and perform nonverbal communications. Accurate recognition of facial expression is essential in many fields, including human-machine interaction, affective computing, robotics, computer games and psychology studies. There are seven basic facial expressions that reflect distinctive psychological activities: anger, disgust, fear, happiness, neutral, sadness and surprise. Examples of these facial expressions are shown in Fig. 1.

This paper focuses on the recognition of two major facial expressions: neutral and smiling. The contribution of the paper is two-fold. First, we propose a method to address localisation error in existing face detection methods. After preliminary face detection, our method applies eye detection, a geometric face model and template-based face verification to precisely locate the face and correct rotation (in-

plane and some out-of-plane rotation). Second, we propose a novel neural architecture for image pattern classification, which consists of fixed and adaptive nonlinear 2-D filters in a hierarchical structure. The fixed filters are used to extract primitive features such as edges, whereas the adaptive filters are trained to extract more complex facial features.

The paper is organized as follows. Section 2 reviews related work on facial expression recognition. Section 3 describes the proposed face detection and alignment method. Section 4 presents the proposed method for recognising smiling and neutral facial expressions. Section 5 analyses the performance of the proposed method on a standard database and compares it with several existing techniques. Section 6 gives the concluding remarks.

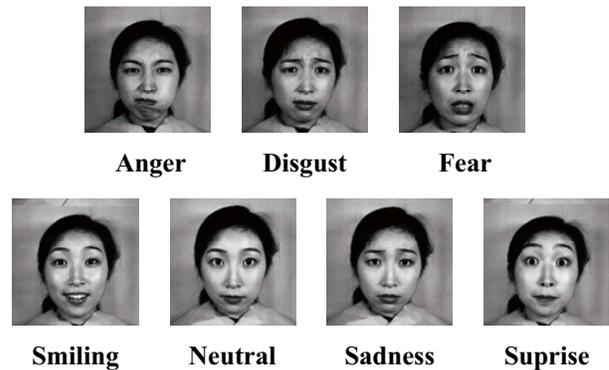


Figure 1. Example facial expressions. The face must be detected, aligned and cropped before facial expression recognition is performed.

## 2. Related Work

Existing approaches for facial expression recognition can be divided into three categories, based on how features are extracted from an image for classification. The categories are geometric-based, appearance-based, and hybrid-based.

## 2.1. Geometric-based approaches

A face image is represented geometrically via fiducial points or the shape of facial regions [7]. Classification is done by analyzing the distances between the fiducial points and the relative sizes of the facial components. Pantic *et al.* [7] proposed a method for detecting facial actions by analyzing the contours of facial components, including the eyes and the mouth. A multi-detector technique is used to spatially sample the contours and detect all facial features. A rule-based classifier is then used to recognize the individual facial muscle action units (AUs). Geometric-based methods cope well with variations in skin patterns or dermatoglyphics. However, they require accurate detection of facial fiducial points, which is difficult when the image has a complex background or a low quality.

## 2.2. Appearance-based approaches

Appearance-based methods process the entire image by applying a set of filters to extract facial features. Zhen *et al.* [15] used Gabor wavelets to represent appearance changes as a set of multi-scale and multi-orientation coefficients. They proposed a ratio-image based feature that is independent of the face albedos. Their method can cope with different people and illumination conditions. Feng [2] used Local Binary Patterns (LBP) to extract facial texture features and combined different local histograms to recover the shape of the face. Classification is performed through a coarse-to-fine scheme, where seven templates were designed to represent the corresponding seven basic facial expressions. At first, two expression classes are selected based on the distance from the test image to the seven templates. The final classification is then done via a K-nearest neighbor classifier with the weighted Chi-square statistic.

## 2.3. Hybrid-based approaches

Facial expression recognition can be improved by combining appearance and geometric features. Zhang and Ji [14] proposed a multi-feature technique that is based on the detection of facial points, nasolabial folds, and edges in the forehead area. In their method, facial features are extracted by associating each AU with a set of movements, and then classified using a Bayesian network model.

## 3. Face Detection and Alignment

Face detection aims to determine the presence and the positions of faces in an image. Many algorithms have been proposed for face detection, e.g. the Viola-Jones face detector [11]. Even with the state-of-the-art face detectors, the boundary for the facial region is sometimes incorrectly calculated. The localisation error is more common when there is rotation (in-plane or out-of-plane). To improve accuracy in facial image analysis tasks such as face, gender

and facial expression recognition, it is essential to correctly align the face image. After initial face detection, we propose a method for face alignment that consists of two main steps. First, for each detected face, candidates for eye points are found using a combination of a Gabor filter and a circular filter. Second, for each eye pair, a face candidate is constructed using a geometric face model, and each face candidate is compared with a face template to remove false detection.

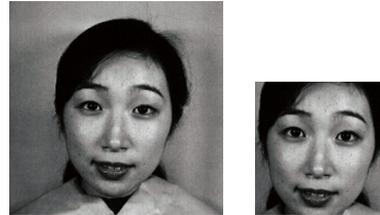


Figure 2. An example of the OpenCV face detector. Even the face is found, its coordinates and rotation angle can be better estimated.

For eye detection, several image preprocessing steps are performed to reduce the effect of the lighting condition. First, the image is histogram equalised. Second, a normalised image is calculated as follows

$$\mathbf{N}(x, y) = \frac{\mathbf{I}(x, y)}{\mu(x, y)}, \quad (1)$$

where  $\mathbf{I}(x, y)$  is an image pixel,  $\mathbf{N}(x, y)$  is the normalized pixel, and  $\mu(x, y)$  is the mean intensity of the neighboring pixels. Third, a contrast-normalization is applied

$$\mathbf{C}_\mathbf{N}(x, y) = \frac{\mathbf{I}(x, y) + \mathbf{N}(x, y)}{\mathbf{I}(x, y) + \mathbf{N}(x, y) + \delta}, \quad (2)$$

where  $\delta$  is the mean pixel intensity of image  $\mathbf{N}$  and the division is performed pixel-wise.

### 3.1. Extracting eye candidates

The candidate eye regions are detected, based on the elongated shape of the eye and the circular shape the the pupil. We will use a Gabor filter and a circular filter. The circular filter was first proposed by Park *et al.* [8].

- Gabor filter

The Gabor filter is the product of a harmonic function and a Gaussian function. The real part of a 2-D Gabor filter is defined as

$$h_1(x, y) = \frac{K_v^2}{\sigma} \exp\left(-\frac{K_v^2(x^2 + y^2)}{2\sigma}\right) * [\cos(x \cdot K_v \cos Q_u + y \cdot K_v \sin Q_u) - \exp\left(-\frac{\sigma}{2}\right)], \quad (3)$$

where  $K_v = \pi \exp(-(v+2)/2)$  and  $Q_u = \pi \times u/6$ . Parameter  $\sigma$  is the ratio of the width of the Gaussian window over the length of the Gabor wavelets. Parameter  $u$  is the orientation of Gabor wavelets. Herein,  $u$  and  $v$  control the orientation and scale of the Gabor wavelets respectively. In this paper, we use  $\sigma = \pi$  and  $Q_u = \pi$ .

- Circular filter

A circular filter is defined as

$$h_2(x, y) = \left( \frac{2}{1 + ((x^2 + y^2)/2\alpha^2)^n} - 1 \right) \times \frac{1}{1 + ((x^2 + y^2)/2\beta^2)^m}, \quad (4)$$

where  $\alpha$  is the inner cutoff variable,  $\beta$  is the outer cutoff variable,  $n$  is the inner order,  $m$  is the outer order,  $\alpha < \beta$ , and  $n < m$ .

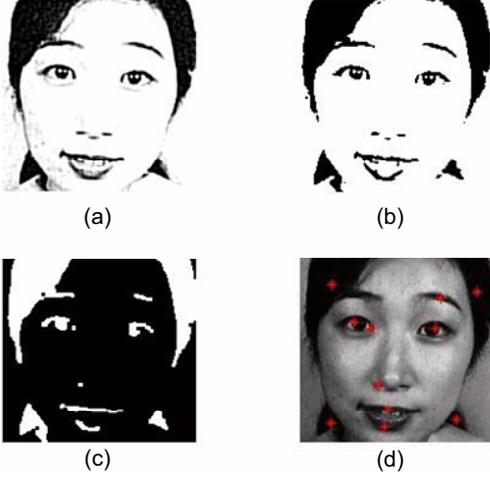


Figure 3. Steps for finding eye candidates: (a) eye filter output, (b) connected component labeling, (c) image erosion and dilation, (d) candidate eye points (shown in red colour).

The Gabor filter and the circular filter are combined to generate the proposed eye filter.

$$h(x, y) = h_1(x, y) + h_2(x, y), \quad (5)$$

The output of the eye filter is shown in Fig. 3a. Then, the connected components are extracted. Image erosion and dilation are performed to remove noise regions. Figure 3d shows eye candidates, after these steps. Note that it is essential to eliminate the false candidates.

### 3.2. Geometric face model and face verification

To eliminate false eye candidates, we construct an image region based on each eye pair and verify if the region is a face pattern. We construct a geometric face model that

reflects the relative anthropometric distances between the facial landmarks, as shown in Fig. 4. This face model is adjustable to out-of-plane rotation [9].

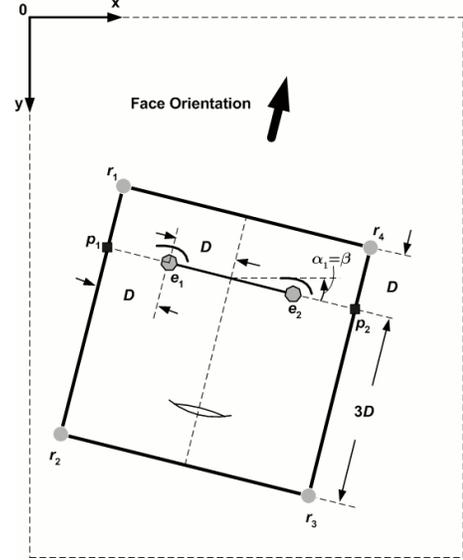


Figure 4. Determining the face boundary for a given eye pair using a geometric face model.

Based on the two eye points  $e_1$  and  $e_2$ , the four corners of the face region are determined as follows.

- Compute the half distance between the two points:

$$D = \frac{1}{2} \sqrt{(e_{2x} - e_{1x})^2 + (e_{2y} - e_{1y})^2}, \quad (6)$$

- Compute the boundary points  $p_1$  and  $p_2$  along the eye line, as in Fig. 4:

$$\begin{aligned} p_{1x} &= (3e_{1x} - e_{2x})/2, & p_{1y} &= (3e_{1y} - e_{2y})/2, \\ p_{2x} &= (3e_{2x} - e_{1x})/2, & p_{2y} &= (3e_{2y} - e_{1y})/2. \end{aligned} \quad (7)$$

- Compute the four face corners  $r_1, r_2, r_3, r_4$ . Assuming that the face angle is  $\alpha$ , the following formulas are used to detection four face corners:

$$\begin{aligned} r_{1x} &= p_{1x} + D \sin \alpha, & r_{1y} &= p_{1y} - D \cos \alpha, \\ r_{2x} &= p_{1x} - 3D \sin \alpha, & r_{2y} &= p_{1y} + 3D \cos \alpha, \\ r_{3x} &= p_{2x} - 3D \sin \alpha, & r_{3y} &= p_{2y} + 3D \cos \alpha, \\ r_{4x} &= p_{2x} + D \sin \alpha, & r_{4y} &= p_{2y} - D \cos \alpha. \end{aligned} \quad (8)$$

- Form face candidates. A face mask can be formed which is then rotated by a face angle  $\alpha$  to the upright position.

- Compare aligned face candidates with a face template. The face template (Fig. 5) used in this paper is generated by averaging 15,000 aligned upright frontal face



Figure 5. A face template constructed using 15,000 frontal face images from the ECU database [9].

patterns. The correlation coefficient between the face candidate and the face template is calculated. Among the overlapping candidates, the one with the maximum correlation score is considered as the true face. Figure 6b shows the corrected upright face image after rotation by the angle of  $\alpha$ .



Figure 6. Eye detection and face alignment: (a) valid eye points after face verification, (b) aligned face pattern.

#### 4. Classification of Smiling and Neutral

The smiling and neutral facial expressions are differentiated using a new neural architecture, which consists of three processing stages as shown in Fig. 7. The first and second stages consist of nonlinear filters, which are used for extracting visual features. The third stage performs classification.

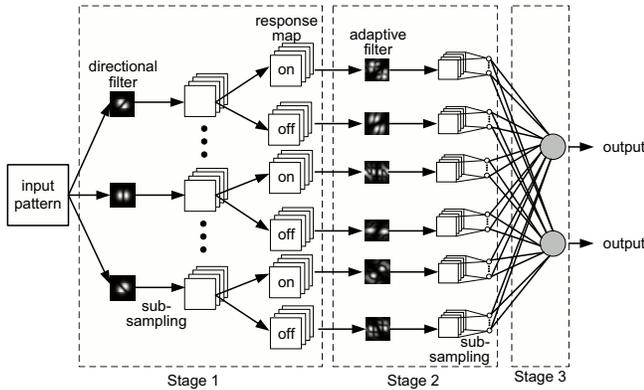


Figure 7. Block diagram of the proposed system.

##### 4.1. Stage 1 - Directional Filters

Stage 1 is designed to extract features at different orientations. It consists of a set of nonlinear filters that are based on a biological mechanism known as *shunting inhibition*. This

mechanism, found in the early visual system [1], has been applied to improve image contrast [4]. The output response of the proposed directional nonlinear filter is computed as

$$\mathbf{Z}_{1,i} = \frac{\mathbf{D}_i * \mathbf{I}}{\mathbf{G} * \mathbf{I}}, \quad (9)$$

where  $\mathbf{I}$  is a 2-D input face pattern,  $\mathbf{Z}_{1,i}$  is the output of the  $i$ -th filter,  $\mathbf{D}_i$  and  $\mathbf{G}$  are the filter coefficients, “\*” denotes 2-D convolution, and the division is done pixel-wise. In this paper, the subscripts 1 and 2 in  $\mathbf{Z}_{1,i}$  and  $\mathbf{Z}_{2,i}$  indicate the outputs of the first and second processing steps, respectively. The kernel  $\mathbf{G}$  is chosen as an isotropic Gaussian kernel:

$$\mathbf{G}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (10)$$

To extract elementary facial features at different directions, the kernel  $\mathbf{D}_i$  is formulated as the  $M$ -th order derivative Gaussian. Its coefficients is defined as

$$\mathbf{D}_i(x, y) = \sum_{k=0}^M \frac{M!}{k!(M-k)!} s_x^k s_y^{M-k} \frac{\partial^M \mathbf{G}(x, y)}{\partial x^k \partial y^{M-k}}, \quad (11)$$

where  $M$  is the derivative order,  $M = 1, 2, \dots$ ,  $\theta_i$  is the angle of rotation,  $\theta_i = \frac{(i-1)\pi}{N_1}$  for  $i = 1, 2, \dots, N_1$ , and  $s_x = \sin \theta_i$  and  $s_y = \cos \theta_i$ .

The partial derivative of the Gaussian with respect to dimension  $x$  or  $y$  can be computed as the product of the Hermite polynomial and the Gaussian function,

$$\frac{\partial^k \mathbf{G}(x, y)}{\partial x^k} = \frac{(-1)^k}{(\sqrt{2}\sigma)^k} H_k\left(\frac{x}{\sqrt{2}\sigma}\right) \mathbf{G}(x, y), \quad (12)$$

where  $H_k(\cdot)$  is the Hermite polynomial of order  $k$ . Figure 8 shows the outputs of directional, derivative Gaussian filters when  $N_1 = 4$  and  $M = 2$ .

Robust image classification requires visual features that are tolerant to small translations and geometric distortions in the input image. To achieve this, we perform a sub-sampling operation and decompose each filter output  $\mathbf{Z}_{1,i}$  into four smaller maps:

$$\mathbf{Z}_{1,i} \rightarrow \{\mathbf{Z}_{2,4i-3}, \mathbf{Z}_{2,4i-2}, \mathbf{Z}_{2,4i-1}, \mathbf{Z}_{2,4i}\}. \quad (13)$$

The first map  $\mathbf{Z}_{2,4i-3}$  is formed from the odd rows and odd columns in  $\mathbf{Z}_{1,i}$ ; the second map  $\mathbf{Z}_{2,4i-2}$  is formed from the odd rows and even columns, and so on.

The next processing step is motivated by the center-surround receptive fields and the two configurations are on-center and off-center. Herein, we separate each sub-sampled map  $\mathbf{Z}_{2,i}$ , where  $i = 1, 2, \dots, 4N_1$ , into an on-response map and an off-response map, using zero as a threshold [10]. Each map is contrast-normalized to produce  $\mathbf{Z}_{4,i}$ .



Figure 8. Outputs of the directional, second-order derivative Gaussian filters for input image 5 of Fig. 1. The parameters are  $N_1 = 4$  and  $\theta_i = 0^\circ, 45^\circ, 90^\circ$  and  $135^\circ$ .

## 4.2. Stage 2 - Trainable Filters

Stage 2 aims to detect more complex features for classification. The output maps produced by each filter in Stage 1 are processed by exactly two filters in Stage 2: one filter for the on-response and the other filter for the off-response. Hence, the number of filters,  $N_2$ , in Stage 2 is twice the number of filters in Stage 1:  $N_2 = 2N_1$ .

Stage 2 is also based on the shunting inhibition mechanism. Consider an input map  $\mathbf{Z}_{4,i}$  to Stage 2. Suppose that  $\mathbf{P}_k$  and  $\mathbf{Q}_k$  are two adaptive kernels for the filter that corresponds to this input map. The filter output is calculated as

$$\mathbf{Z}_{5,i} = \frac{g(\mathbf{P}_k * \mathbf{Z}_{4,i} + b_k) + c_k}{a_k + f(\mathbf{Q}_k * \mathbf{Z}_{4,i} + d_k)}, \quad (14)$$

where  $a_k, b_k, c_k$  and  $d_k$  are adjustable bias terms, and  $f$  and  $g$  are two activation functions. A sub-sampling operation is performed across each set of four output maps generated from the adaptive filter by averaging each non-overlapping block of size  $(2 \times 2 \text{ pixels}) \times (4 \text{ maps})$  into a single output signal:

$$\{\mathbf{Z}_{5,4i-3}, \mathbf{Z}_{5,4i-2}, \mathbf{Z}_{5,4i-1}, \mathbf{Z}_{5,4i}\} \rightarrow \mathbf{Z}_{6,i}. \quad (15)$$

This sub-sampling process is repeated for each adaptive filter to generate a feature vector.

## 4.3. Stage 3 - Classification

The extracted features are sent to Stage 3 for classification. Stage 3 may use any type of classifiers. Previously, we used a linear classifier whose output  $y_j$  is given as

$$y_j = \sum_{i=1}^{N_3} w_{i,j} \mathbf{Z}_{6,i} + b_j, \quad j = 1, 2, \dots, N_4 \quad (16)$$

where  $w_{i,j}$ 's are adjustable weights,  $b_j$  is an adjustable bias term,  $\mathbf{Z}_{6,i}$ 's are input features to Stage 3,  $N_3$  is the number of input features, and  $N_4$  is the number of output nodes. The output  $\mathbf{y} = [y_1, y_2, \dots, y_{N_4}]^T$  indicates the class or the label of the input pattern  $\mathbf{I}$ .

We use a second-order training method called the Levenberg-Marquardt (LM) algorithm [3] to optimize the parameters of the adaptive filters in Stage 2 and the classifier in Stage 3.

## 5. Results and Analysis

We analyse the performance of the proposed method on the standard Japanese Female Facial Expression (JAFFE) database [6], which is commonly used in research on facial expression recognition. This database consists of 213 images from 10 Japanese actresses. They were instructed to produce seven types of facial expressions (see Fig. 1). For each person, two to four images were recorded for each facial expression.

We applied the 10-fold cross validation on the *smile* and *neutral* expressions of the JAFFE database. All images were divided into ten groups. For each validation fold, nine groups were used to train the classifier while the remaining group was used for testing. This step was repeated 10 times, and the classification rates of the ten folds were averaged to form the final estimate of the classification rate. The proposed system uses an input image size of  $42 \times 42$  pixels. Experiments were conducted to determine the suitable values for different parameters [5]. Stage 1 uses the second-order Gaussian derivative ( $M = 2$ ) and four directions ( $N_1 = 4$ ). The filter sizes for Stages 1 and 2 are 7-by-7 and 3-by-3 pixels, respectively.

We also used JAFFE database to evaluate the proposed eye detection and face alignment method. As shown in Table 1, the detection rate of the proposed method is comparable to other existing methods, tested on the same database.

Table 1. Eye detection performance on the JAFFE database.

Method	Year	Detection rate (%)
Proposed method	2010	100.0
Wang <i>et al.</i> [13]	2008	99.1
Wang and Yin [12]	2005	95.8
Zhou and Geng [16]	2004	97.2

Experiments were conducted to evaluate the proposed method for differentiating smiling and neutral facial expressions. Results in Table 2 indicate that recognition performance increases when the face pattern is better aligned and cropped and in-plane rotation is corrected. Table 3 shows the confusion matrix for smiling and neutral recognition, when the face image is aligned. The overall recognition accuracy is 99.00%. This method can also cope with images containing multiple faces. For each face detected by OpenCV, we apply the proposed face alignment method. The corresponding facial expression is then determined by passing the upright face image through the FER system. Figure 9 gives a visual example of the overall automatic FER system.

Table 2. Comparison of recognition accuracy for non-aligned and aligned faces on the JAFFE database.

Method	CR (%)
Non-aligned faces (OpenCV)	98.67
Aligned faces (proposed)	99.00

Table 3. Confusion matrix for the two facial expressions. The entry at (row  $r$ , column  $c$ ) is the percentage of facial expression  $r$  that is classified as facial expression  $c$ .

%	Smiling	Neutral
Smiling	98.00	2.00
Neutral	0.00	100.00

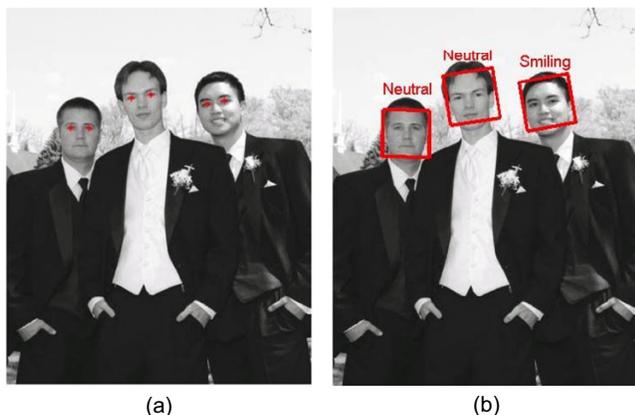


Figure 9. A visual result of the automatic system: (a) eye detection, (b) expression recognition.

## 6. CONCLUSION

We presented an approach for automatic recognition of smiling and neutral faces. In our approach, eye detection and face alignment are used to correct localisation errors in the existing face detectors. The classification between smiling and neutral faces is done via a novel neural architecture that combines fixed, directional filters and adaptive filters in cascade. The fixed, directional filters extract primitive edge features, whereas the adaptive filters are trained to extract more complex features, which is classified by the linear classifier. The eye detection and face alignment method can correctly process all images in the JAFFE database. On this benchmark database, our system achieves a classification rate of 99.0% for smiling and neutral faces.

## References

- [1] L. J. Borg-Graham, C. Monier, and Y. Fregnac. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 393(6683):369–373, 1998. 4
- [2] X. Feng. Facial expression recognition based on local binary patterns and coarse-to-fine classification. In *Proceedings of*

*Fourth International Conference on Computer and Information Technology*, pages 178–183, 2004. 2

- [3] M. T. Hagan and M. B. Menhaj. Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6):989–993, 1994. 5
- [4] T. Hammadou and A. Bouzerdom. Novel image enhancement technique using shunting inhibitory cellular neural networks. *IEEE Transactions on Consumer Electronics*, 47(4):934–940, 2001. 4
- [5] P. Li, S. L. Phung, A. Bouzerdom, and F. H. C. Tivive. Improved facial expression recognition with trainable 2-d filters and support vector machines. In *Proceedings of the 20th International Conference on Pattern Recognition*, 2010. 5
- [6] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998. 5
- [7] M. Pantic and L. J. M. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(3):1449–1461, 2004. 2
- [8] C. W. Park, K. T. Park, and Y. S. Moon. Eye detection using eye filter and minimisation of nmf-based reconstruction error in facial image. *Electronics Letters*, 46(2):130–132, 2010. 2
- [9] S. L. Phung and A. Bouzerdom. A pyramidal neural network for visual pattern recognition. *IEEE Transactions on Neural Networks*, 18(2):329–343, 2007. 3, 4
- [10] F. H. C. Tivive, A. Bouzerdom, S. L. P. Phung, and K. M. Liftekharuddin. Adaptive hierarchical architecture for visual recognition. *Applied Optics*, 49:B1–B8, 2010. 4
- [11] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2004. 2
- [12] J. Wang and L. Yin. Detecting and tracking eyes through dynamic terrain feature matching. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 4, pages 78–85, 2005. 5
- [13] Q. Wang, C. Zhao, and J. Yang. Efficient facial feature detection using entropy and svm. In *ISVC '08: Proceedings of the 4th International Symposium on Advances in Visual Computing*, pages 763–771. Springer-Verlag, 2008. 5
- [14] Y. Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):699–714, 2005. 2
- [15] W. Zhen and T. S. Huang. Capturing subtle facial motions in 3d face tracking. In *Proceedings of Ninth IEEE International Conference on Computer Vision*, pages 1343–1350, 2003. 2
- [16] Z. H. Zhou and X. Geng. Projection function for eye detection. *International Journal of Pattern Recognition*, 37:1049–1056, 2004. 5