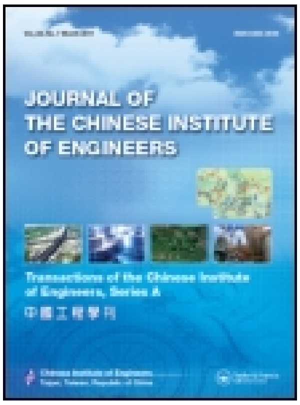


Journal of the Chinese Institute of Engineers

Publication details, including instructions for authors and subscription information: http://www.tandfonline.com/loi/tcie20

Automated facial expression recognition system using neural networks

Jyh-Yeong Chang [a] & Jia-Lin Chen [a]

[a] Department of Electrical and Control Engineering , National Chiao Tung University , Hsinchu, Taiwan 300, R.O.C.

Published online: 03 Mar 2011.

To cite this article: Jyh-Yeong Chang & Jia-Lin Chen (2001) Automated facial expression recognition system using neural networks, Journal of the Chinese Institute of Engineers, 24:3, 345-356, DOI: 10.1080/02533839.2001.9670632

To link to this article: http://dx.doi.org/10.1080/02533839.2001.9670632

AUTOMATED FACIAL EXPRESSION RECOGNITION SYSTEM USING NEURAL NETWORKS

Jyh-Yeong Chang* and Jia-Lin Chen
Department of Electrical and Control Engineering
National Chiao Tung University
Hsinchu, Taiwan 300, R.O.C.

Key Words: facial expression recognition, neural classifier, point contour detection method, facial action unit.

ABSTRACT

This paper proposes an automated facial expression recognition system using neural network classifiers. First, we use the rough contour estimation routine, mathematical morphology, and point contour detection method to extract the precise contours of the eyebrows, eyes, and mouth of a face image. Then we define 30 facial characteristic points to describe the position and shape of these three facial features. Facial expressions can be described by combining different action units, which are specified by the basic muscle movements of a human face. We choose six main action units, composed of facial characteristic point movements, as the input vectors of two different neural network-based expression classifiers including a radial basis function network and a multilayer perceptron network. Using these two networks, we have obtained recognition rates as high as 92.1% in categorizing the facial expressions neutral, anger, or happiness. Simulation results by the computer demonstrate that computers are capable of extracting high-level or abstract information like humans.

I. INTRODUCTION

As personal computers advance day by day, computers become necessities in our daily life. Computers are of great help in many areas, such as engineering, commerce, entertainment, and so forth. Today, we hope that computers can not only execute various computation-intensive algorithmic routines but can also extract high-level or vague information for us. For example, if the computer knows human emotion, then an automated facial expression recognition system could create a new era in providing a natural and friendly interface between human and machine. If an abstract and uncertain message, facial expression for instance, can be recognized by the computer, other kinds of abstract and vague information can also be recognized by a computer. This breakthrough will greatly enlarge the application domains of the computer. In this perspective, this paper is concerned with investigating an automated facial expression recognition system so that the computer can understand the emotional messages conveyed from one's facial expression.

Aside from the intelligent human computer interface described above, facial expressions and their recognition also provide an important behavior measure for the study of emotion, cognition process, and social interaction (Cohn *et al*., 1999). Scientific study of facial expressions began with the team led by Ekman (Ekman and Friesen, 1975, 1978). They

*Correspondence addressee

analyzed six facial expressions, which included surprise, fear, disgust, anger, happiness, and sadness. Each expression was summarized by distinctive clues in the appearance of the eyebrows, eyes, mouth, jaw, etc. These facial expression clues are further investigated and encoded into the so-called Facial Action Coding System (FACS) (Ekman and Friesen, 1978), to describe "all visually distinguishable facial movements." FACS enumerates "Action Units" (AUs) of a face that cause facial movements. In FACS, there are 46 AUs that account for changes in facial expression. The combination of these action units results in a large set of possible facial expressions.

In this decade, many methods for recognizing facial expressions by machines have been presented. Among these approaches, four types can be categorized. The first type uses emotion space to recognize facial expression (Morishima and Harashima, 1993). The emotion space model of facial expressions was built from training face images with different expressions. A face expression is recognized as the one with the largest projected component in the emotion space, similar to the eigenface approach used for face recognition (Sirovich and Kirby, 1987; Turk and Pentland, 1991a,b). The second type is to recognize facial expressions of an image frame sequence by the use of optical flow (Mase, 1991; Rosenblum *et al*., 1996). Facial expressions are the result of facial muscle actions which are triggered by the nerve impulses generated by emotions. The muscle actions cause the movement and deformation of facial skin and facial features such as eyes, mouth, and nose. We can use optical flow to estimate facial muscle actions which can then be used in recognizing a facial expression. The third type is to use active shape models to recognize facial expressions (Cootes *et al*., 1995; Lanitis *et al*., 1997). This addresses facial expression recognition by establishing the distribution of appearance parameters over a selected training set for each expression category so that the appearance parameters calculated for a new face image can be determined and then used for expression classification. The fourth type is to recognize facial expressions by neural network (Kobayashi and Hara, 1992, 1994; Matsuno *et al*., 1995; Padgett and Cottrell, 1997), namely, under a framework exploiting neural models to capture and encode the nonlinear mapping among different facial expressions. Among these four approaches, the optical flow approach classifies facial expression from image sequences and the other three expression recognition schemes from a static image. Expression recognition from static images is much more difficult than from image sequences because only subtle and vague information is disclosed in a single image from one expression to another. In the neural network-based approaches above (Kobayashi and Hara, 1992, 1994), the technique involving facial features and fiducial points extraction is not addressed; they are extracted manually. To circumvent this shortcoming, we have developed an efficient facial feature extraction scheme to build a fully automated expression recognition system. Instead of using 60 inputs of *x* and *y* coordinates obtained from 30 facial fiducial points directly, a sound selection of only six efficient AUs as neural classifier inputs results in a great simplification in the neural structure being used, 6×10×3 in our system in comparison with 60×100×100×6 in Kobayashi and Hara's (1992). Aside from such great simplification in structure, we have still obtained a recognition rate as high as 92.1%, which outperforms their approach by 2%.

This paper proposes an automated facial expression recognition system using neural models. First, we extract, by computer vision techniques, three main facial features: eyebrows, eyes, and mouth of a front-view face image. We use Rough Contour Estimation Routine (RCER) to obtain the contours of eyebrows, eyes, and mouth (Chen, 1991; Huang and Chen, 1992). Because the eyebrow contours obtained by RCER were precise enough for facial expression recognition, they were not further refined. However, the shapes of eyes and mouth are not precise enough, so we refined them by a newly developed scheme, called Point Contour Detection Method (PCDM), to improve the precision of the contours of the eyes and mouth. After extracting these features' contours, we defined 30 facial characteristic points to describe the position and shape of the features. In the area of psychological research, action units are used for describing the basic muscle movement of the human face (Ekman and Friesen, 1978). Facial expressions can thus be described by combining different action units. We chose six dominant action units, composed of facial characteristic point movements, as the inputs for the neural network-based expression classifier. The proposed neural system recognizes the expressions based on facial characteristic point (FCP) changes shown in facial features of the eyebrows, eyes, and mouth. However, many sources of variability, including the facial skin brightness, illumination conditions, 3D pose, and individual appearance, make it difficult to estimate precise positions and shape attributes of face features in a real image. Expression recognition, particularly from static images, is even more difficult due to the lack of rich evidence of correlation exhibited in the timing axis. Even we as human observers may make mistakes and often fail to agree in categorizing an expression from a single image (Lanitis *et al*., 1997; Zhang, 1997). This source of disagreement may account for around 15-20% of the error in labeling seven, the previously-noted six expressions and neutral as well, expressions over the testing image instances. As a practical matter, we restrict our recognition to the most common expressions, neutral, anger, and happiness, seen in our daily life. Consequently, the variability of human disagreement in labeling an expression can be avoided, which in turn improves the correctness of our experimental validation phase.

The rest of this paper is organized as follows. In Section II, feature extraction modules for facial expression recognition are introduced. First, we briefly outline the rough contour estimation routine (RCER) to obtain the boundaries of the facial features including eyebrow, eye, and mouth. In the sequel, the PCDM technique for precise contour estimation of the eyes and the mouth is introduced. Facial characteristic points of a face are then introduced in Section. III. We exploit six action units to encode movements of facial muscles in terms of the positional changes of relevant FCPs. Then two facial expression classifiers, which are implemented by a radial basis function network (RBFN) and a multilayer perceptron (MLP) network, are proposed using these six AUs as inputs. The computer simulations and their results are presented in Section IV. Finally, the conclusion is given in Section V.

II. FEATURE EXTRACTION MODULES FOR FACIAL EXPRESSION RECOGNITION

The human face has several features, eyebrows, eyes, mouth, nose, which together with the facial outline, are suitable for identification or expression recognition. Recent advances have been made (Cootes *et al.*, 1995; Shackleton and Welsh, 1991; Yuille *et al.*, 1992) in extracting facial feature for recognizing faces and facial expressions. In this contribution, we extract three main facial features: eyebrows, eyes, and mouth for facial expression labeling. After extracting these features, we are able to locate 30 Facial Characteristic Points (FCPs). FCPs are a set of fiducial points defined on these facial features and hence carry information about the position and shape of these three features. According to Ekman and Friesen (1975), almost all expressions of the human face are described by the combination of 46 basic movements of facial muscles and these basic movements are called Action Units (AUs). In this paper, the method for defining AUs through calculating positional changes of FCPs is the same as the approach proposed by Kobayashi and Hara (1992, 1994).

1. Rough Contour Estimation Routine

To begin with, we will introduce an algorithm to estimate the rough contour of the facial feature. The facial feature contour is estimated by the rough contour estimation routine, which is a pixel-wise region growing algorithm based on gray level. The goal of this algorithm is to locate a connected region containing the same object. We use this algorithm to obtain the contours of the facial features which include eyebrow, eye, and mouth. In summary, the RCER algorithm consists of the following two steps:

(i) Search for a connected region and find the maximum and minimum *y*-coordinates corresponding to each *x*-coordinate belonging to the connected region.

(ii)Combine the maximum and minimum *y*-coordinates in each *x*-coordinate belonging to the connected region to get the rough contour.

The interested reader may refer to (Chen, 1991; Huang and Chen, 1992) for details about the RCER algorithm.

2. Point Contour Detection Method for Facial Contour Estimation

After the rough contour is obtained by the RCER, refining an accurate contour for each facial feature is the next step. Because eyebrows are absolutely black compared with other nearby regions and the contour of the eyebrows is relatively simple, the eyebrow contour obtained by RCER is precise enough for facial expression recognition, viz., no further refinement of eyebrow contour is needed. On the other hand, we will refine the precise contours of eyes and mouth by a newly-developed scheme, called Point Contour Detection Method (PCDM), which will be described in the following.

Extracting the boundaries of facial features is difficult for the following two reasons. First, the edges of facial features are usually not organized into a sensible global percept. Second, the low gray level contrasts around facial features cause the boundary detection difficulties. Consequently, the detection of facial features by conventional edge detection routines is often not successful and one usually resorts to new edge detection schemes. Among these, Active Shape Model (ASM) (Cootes *et al.*, 1995) and Deformable Template Model (DTM) (Shackleton and Welsh, 1991; Yuille *et al.*, 1989, 1992) are most frequently cited in the literature. ASM, which uses Point Distribution Model (PDM) in the image search is a compact parameterized model to outline facial features. The model represents both shape and gray-level appearance and is created by performing a statistical analysis over a training set of face images. The PDM is quite powerful in describing various kinds of contours, but it is very computationally involved. In the DTM method, the deformable templates are specified by a set of parameters which use *a priori* knowledge about the expected shape of the features to guide the contour deformation process. The templates are flexible enough to be able to change their size and other parameter values to match themselves to the data. An energy function is defined and then minimized which contains terms attracting the template to salient features according to image and edge intensity. The minimum obtained corresponds to the best fit of the deforming contour with the target feature in the image. However, the method is also very complex and computationally expensive in finding the minimum of the energy function. This heavy computational load becomes a restriction to a real-time facial expression recognition system whenever required. Here, we will propose the point contour detection method to improve the contours of the eyes and mouth. Our method uses the advantages of the two schemes stated above. For the sake of computational efficiency, we first borrow the landmark points concept of ASM to describe the contour. Second, we stick to the spirit of DTM in relocating the contour, i.e., the landmark points of the feature's boundary are updated by searching around the strongest edge strength. The proposed PCDM eliminates rather complex routines in deriving point distribution models and speeds up the updating of the landmark points, which requires lengthy computation in the ASM scheme. Our PCDM is reliable and much simpler in identifying the eyes and the mouth automatically. The details of the PCDM are illustrated in the following.

(i) The Point Contour Detection Method

Before performing feature extraction, we generate the edge image by the morphology technique. We use two operations, dilation and erosion, to obtain the potential field for the edge intensity. The dilation and erosion operators for gray-scale images are defined conventionally (Gonzalez and Woods, 1992). We define the edge intensity image ϕ_{edge} by

$$\phi_{edge}=dilation-erosion. \tag{1}$$

Contours of the salient features, such as eyebrows, eyes, mouth, and nose, will demonstrate stronger intensity on the edge intensity image ϕ_{edge}. Using ϕ_{edge}, the original image, and *a priori* knowledge of a face, we can locate a target facial feature so that it is confined in a block. When the block is obtained, we generate some landmark points which are marked at equal distances on the upper and lower boundaries of the block. Such landmark points are used for the initial contour points of the target feature. Because the target feature is inside the block and the edge strength on the contour of the target feature is stronger, a learning scheme was developed to force the landmark points on the upper boundary to search downward and the points on the lower boundary to search upward for best matching the landmark points with the local edge. After the local edge is specified by the landmark points via learning, these landmark points can be connected to form a closed curve to outline the contour. Because of variability exhibited in a facial image, some landmark points obtained by the present PCDM can subsume the wrong positions and deviate from the true contour. Hence, the closed curve connecting these contour landmark points directly could not be smooth enough. Since the contours of the eyes and mouth are analytically describable, it is advantageous to use the pseudo-inverse technique to find a smooth quadratic curve that connects these landmark points with a minimum square error in the distance.

(ii) Eyes Extraction

According to the observation reported in Huang and Chen (1992), the position of the left eyebrow is about one-fourth of the facial width. With this *a priori* knowledge in mind and observing the intensity of the edge image ϕ_{edge} and the original gray level image simultaneously, we can easily find the left eye and confine it by a rectangular block. As noted above, the eyebrows can be extracted precisely by the RCER routine, therefore the length of the left eyebrow is known. Because the left eye is usually shorter than the left eyebrow and below the left eyebrow, we can easily specify a block to include the left eye. The width of the block is fixed with enough tolerance to include the left eye of all face images. The length of the block, however, is an adjustable parameter specified by the corner of the eye and hence varies from instance to instance. For the edge image ϕ_{edge} obtained from morphology, we found that the intensity of the pixels outside the eye usually is smaller than a preset threshold while the pixel intensity on the eye contour is usually greater than the threshold. In our experience, a value of 75 for the threshold is acceptable for our training and test sets of images. The largest and smallest horizontal positions of the pixels greater than this threshold could be chosen as the two corners of the left eye. These two corner points will be used to define the length of the initial block that confines the left eye. In this block, we then, respectively, generate five landmark points at equal distances on the upper and lower boundaries as shown in Fig. 1(a). Since the edge strength on the contour of the eye is stronger, each landmark point on the upper boundary will sequentially move downward gradually to be located at a point where the edge

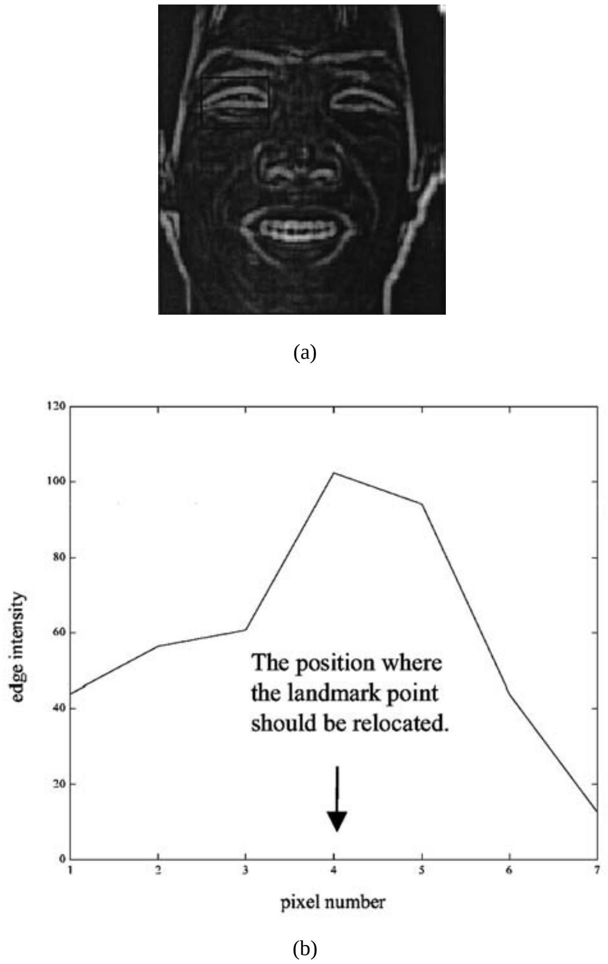


Fig. 1 (a) The initial landmark points (denoted by ■) defined on the confined block of the right eye; (b) The edge intensity versus vertical positions of a typical landmark point

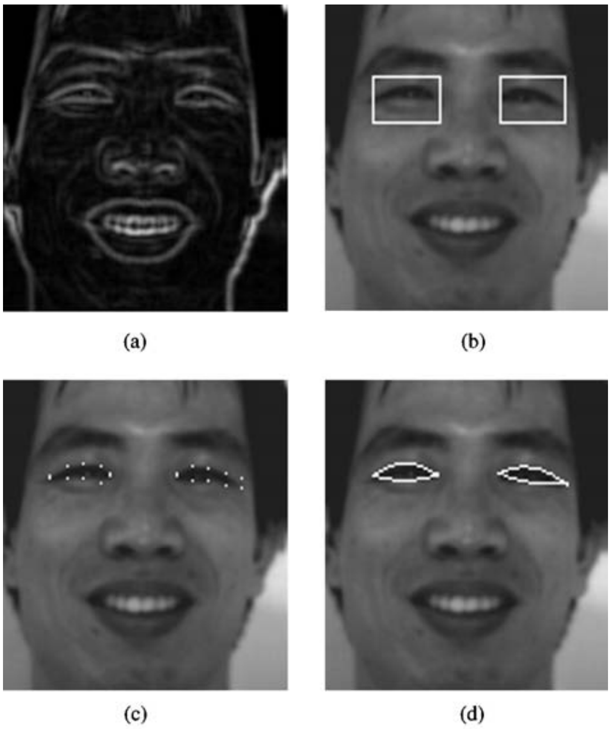


Fig. 2 An eye extraction procedure. (a) A morphology-based edge image; (b) Initial rectangular blocks to confine the eyes; (c) Final landmark points for eye contours; (d) Extracted eye contours.

intensity is maximal in the block. A typical landmark point relocating process of the above is demonstrated in Fig. 1(b). After all the upper landmark points are moved to the edge, we then derive a quadratic curve to connect these five upper landmark points for minimum square error in the distance using the pseudo-inverse solution. Similarly, we can move the lower five landmark points upward gradually to the lower edge of the eye and use another quadratic curve to connect these lower landmark points with the minimal square error in the distance. A typical eye extraction procedure, which includes several intermediate output images, is shown in Fig. 2. Note that the final landmark points shown in Fig. 2(c) are denoted by white spot symbols, while the initial landmark points of Fig. 1(a) are denoted by black spot symbols subject to a better contrast consideration.

(iii) Mouth Extraction

The procedure for mouth extraction is similar to that for eye extraction. Because the images are front-views of human faces, the mouth should be located in a position somewhere between the eyes and lower. We can go downward from the middle of the left and right eye to locate the approximate position of the mouth. According to our observation and referring to (Huang and Chen, 1992), the height of the higher lip is between one-tenth and one-third of the mouth's length, whereas the height of the lower lip is between one-sixth and two-fifths of the mouth's length. As with eye extraction, it is appropriate, for our training and test sets of images, to set the threshold to be 70 to specify the largest and smallest horizontal positions, i.e., length, of the mouth block. With this *a priori* knowledge and observing the edge image ϕ_{edge}, we can define a block, in a manner similar to the eyes, so that the mouth is inside this block. For all of the training and testing images, we have found that the edge strength on the contour of the mouth is stronger also. Then we choose nine landmark points at equal distances on the upper and lower boundaries of the block, respectively. Each of the nine landmark points on the upper boundary will move downward gradually to the position which gives the maximum intensity. The lower nine landmark points will be relocated in a similar manner. When all the landmark points are adjusted to the contour of the mouth, we derive two quadratic equations,

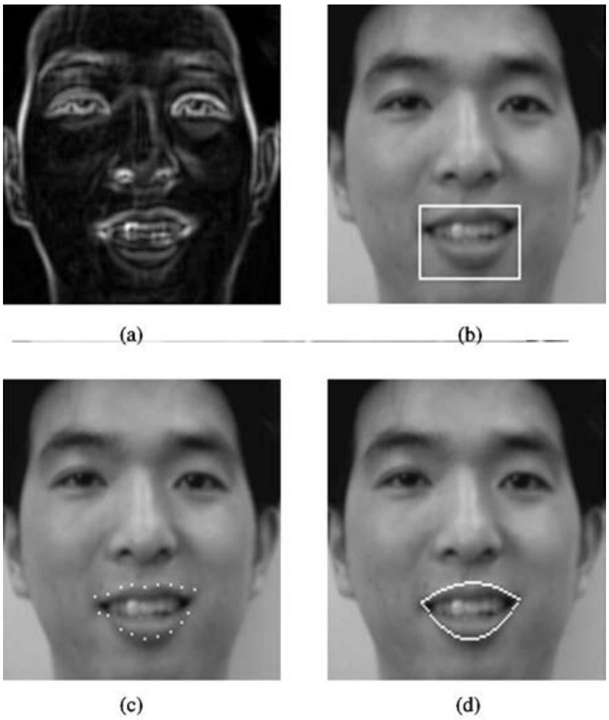


Fig. 3 A mouth extraction procedure. (a) A morphology-based edge image; (b) Initial rectangular block to confine the mouth; (c) Final landmark points for mouth contours; (d) Extracted mouth contours.

respectively, that connect the upper and lower landmark points by the pseudo-inverse method. A typical mouth extraction procedure, which includes several intermediate output images, is shown in Fig. 3.

III. USING ACTION UNITS TO RECOGNIZE FACIAL EXPRESSIONS

In this section, 30 facial characteristic points will be defined for the three facial features of eyebrows, eyes, and mouth. These facial characteristic points are composed to define action units. Finally, we recognize facial expressions through action unit matching using neural network models. The details are described in the following.

1. Facial Characteristic Points

Facial characteristic points are the fiducial points in a face which represent facial characteristics (Kobayashi and Hara, 1994). Fig. 4 shows the facial characteristic points, $a_i' s$, a_i is a vector defining the coordinate of the i-th FCP, i.e., a_i is described as

$$a_i=(x_{ai}, y_{ai}), \quad i=1, 2, .., 30. \tag{2}$$

As shown in Fig. 5, the x_a–y_a coordinate system is the *absolute* coordinate system, with its origin being chosen at the lower left corner of an input image, and x''–y'' is the new *reference* coordinate system used for processing FCPs. The origin (O_x, O_y) of the x''–y'' coordinate system is chosen as a point on the length, *base*, downward the mid point between the left and right eyes, which is almost the same center used for image alignment in Cohn *et al*. (1999). Parameter *base* denotes the distance between the two eyes and is given by

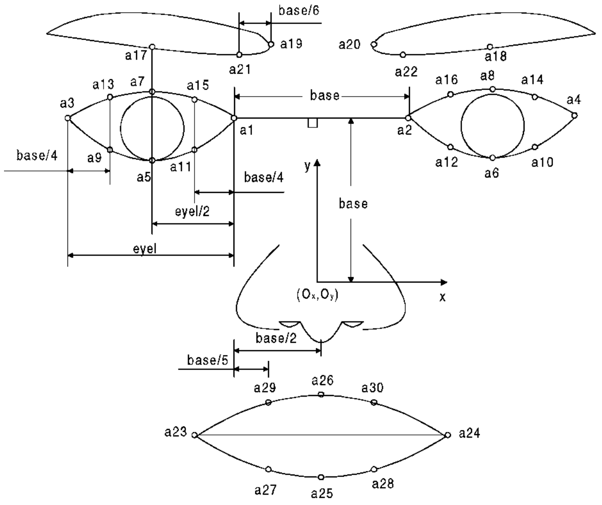


Fig. 4 The facial characteristic points

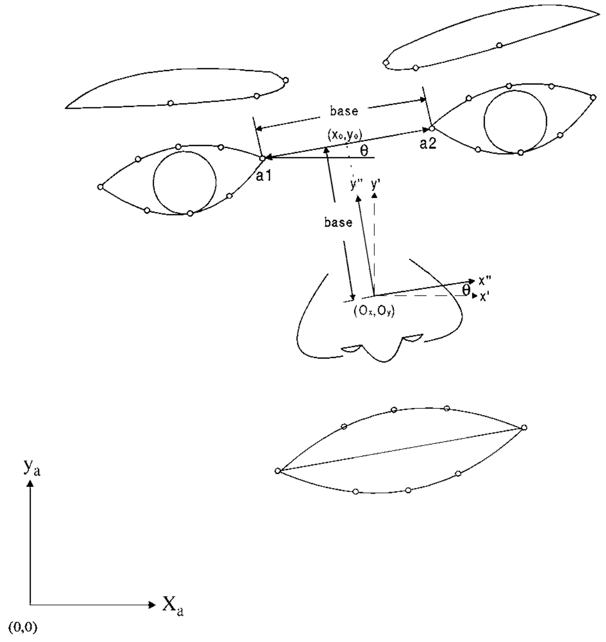


Fig. 5 The absolute and relative coordinate systems

$$base = \sqrt{(x_{a2}-x_{a1})^2+(y_{a2}-y_{a1})^2}\ . \tag{3}$$

It is a length normalization factor so that each image length will be normalized by *base* to make every face image instance have a length of unity between the two eyes. This normalization factor is also employed in Zhang (1999). Furthermore, to offset the 3-D pose of the head in a picture, we introduce a coordinate rotation angle θ, which is the inclination of the face and is specified by the line of the two eyes with respect to the horizontal, defined by

$$\theta = \tan^{-1}\frac{y_{a2}-y_{a1}}{x_{a2}-x_{a1}}\,. \tag{4}$$

Let (x_0, y_0) be the coordinate of the mid point between the left and right eyes, namely,

$$x_0 = \frac{x_{a1}+x_{a2}}{2}\,,$$

$$y_0 = \frac{y_{a1}+y_{a2}}{2}\,. \tag{5}$$

From Fig. 5, the origin of the new coordinate system is then calculated as

$$O_x = x_0 + base \times \sin\theta,$$

$$O_y = y_0 - base \times \cos\theta. \tag{6}$$

The coordinate (x_{ai}, y_{ai}) of an FCP a_i is transformed into the x''–y'' coordinate system, first by translating the origin to the coordinates of Eq. (6) above and then rotating an angle of orientation θ. The relationships to implement this affine transform are given by successively executing the following two sets of equations:

$$x'_{ai} = x_{ai} - O_x\,,$$

$$y'_{ai} = y_{ai} - O_y\,; \tag{7}$$

and

$$x''_{ai} = x'_{ai}\cos\theta + y'_{ai}\sin\theta\,,$$

$$y''_{ai} = -x'_{ai}\sin\theta + y'_{ai}\cos\theta\,. \tag{8}$$

Finally to normalize the input face image, dividing (x''_{ai}, y''_{ai}) by *base* gives

$$\overline{x}_{ai} = \frac{x''_{ai}}{base}\,,$$

$$\overline{y}_{ai} = \frac{y''_{ai}}{base}\,;\quad i=1, 2, ..., 30. \tag{9}$$

After this division, the size of each input face image will become identical in a sense of having unified length between eyes, which can also compensate for the distance effect between a client face and the CCD camera during taking a picture.

Table 1 Action Units for expressing three facial expressions

No.	Appearance Changes
AU 4	Brow lower
AU 6	Cheek raise
AU 7	Lids tight
AU 10	Upper lip raise
AU 12	Lips corner pull
AU 26	Jaw drop

2. Composition of Action Units from Facial Characteristic Points

According to the study of Ekman and Friesen (1978), almost all facial expressions can be described by the combination of 46 AUs. The use of AUs for expression recognition was motivated by the study of Kobayashi and Hara (1994). Continuing the result reported in Ekman and Friesen (1978), they further analyzed the importance of these AUs which cause an expression. Contributing AU components for an expression were identified and, moreover, the factors of their relative importance were also analyzed. Accordingly, we chose six most effective AUs to be used to recognize the following three expressions: neutral, anger, and happiness. Table 1 shows these six AUs, where the way of composing AUs with FCP positions is the same as Kobayashi and Hara (1992, 1994). The facial expression messages are revealed from the (strength) values of AUs, which in turn are represented by the FCP displacements of input face image from the corresponding neuter one.

In Figs. 4 and 5, FCPs are described by a_1 to a_{30} and we assign Arabic number 1 as the symbol to represent the $\overline{x}$-coordinate of a_1 and number 2 as the $\overline{y}$-coordinate of a_1 based on Eqs. (7)-(9). We assign the numerals to other FCPs in a similar manner. For instance, FCP a_{26} will be symbolized by coordinate vector (51, 52). Fig. 6 shows the classification tree structure for AU calculation needed for facial expression recognition. The values of the AUs employed are computed by the following eight computational formulas:

(f-1) $(37_e-37_n)-(39_e-39_n)$

(f-2) $(14_e-14_n)+(16_e-16_n)+(26_e-26_n)+(28_e-28_n)+(30_e-30_n)+(32_e-32_n)$

(f-3) $(10_e-10_n)+(12_e-12_n)+(18_e-18_n)+(20_e-20_n)+(22_e-22_n)+(24_e-24_n)$

(f-4) $(34_e-34_n)+(36_e-36_n)+(38_e-38_n)+(40_e-40_n)+(42_e-42_n)+(44_e-44_n)$

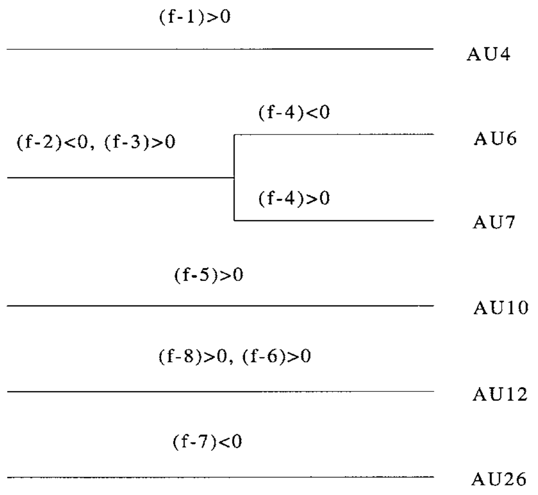


Fig. 6 The classification tree structure for AU calculation

(f-5) $(52_e-52_n)+(58_e-58_n)+(60_e-60_n)$

(f-6) $(46_e-46_n)+(48_e-48_n)$

(f-7) $(50_e-50_n)+(54_e-54_n)+(56_e-56_n)$

(f-8) $-(45_e-45_n)+(47_e-47_n)$

These formulas are computed by using the values corresponding to the facial information numbers from 1 to 60 noted above, where subscript *e* represents the expressive input image and subscript *n* represents the prototypical neutral expression of the corresponding input client face assumed known or recognized beforehand. The values of AUs are finally computed by corresponding computational values defined in the tree structure of Fig. 6. For example, when the values of Formulas (f-2) and (f-3) are, respectively, smaller and larger than zero and, if Formula (f-4) is smaller than zero, then AU6 is the value computed from Formula (f-4), otherwise, the value of AU6 is set to zero.

3. Recognition of Facial Expressions by Neural Classifiers

After obtaining the AUs, we are able to recognize the facial expressions by a classification scheme. The information processing regarding facial expression recognition is somewhat abstract and high level and there exists a great deal of uncertainty, vagueness, and imprecision in various phases of the information processing. A classifier which can accommodate such imprecision and uncertainty, such as a neural model, may be more appropriate for this purpose. Neural network-based classifiers are designed in a way to mimic the human-like mechanism to perform a classification task and has demonstrated many successful examples in an unknown and/or unstructured environment. In particular, radial basis function network (RBFN) (Howell and Buxton, 1995; Rosenblum *et al.*, 1996) and multilayer perceptron (MLP) neural network (Kobayashi and Hara, 1994; Zhang, 1999) are most widely exploited to perform classification tasks on human faces. We have also embedded these two connectionist models in our facial expression recognition system, whose details will be illustrated in the following section.

IV . SIMULATION

1. Database of The Face Images

The front-view face images were acquired using a CCD camera, under almost identical environments of distance, illumination, and background. The resolution of images is 128×128 pixels. There were 80 face images collected in our face image database, which includes eight different male persons and each person poses for several neutral, anger, and happiness expressions. Examples of expressive images collected in the database are shown in Fig. 7. In our database, we used 42 images as training data and 38 images for testing, both of which include neutral, anger, and happiness expressions.

2. Recognition Results

First, we used RCER to obtain the rough contours of eyebrows, eyes, and mouth. Because the eyebrow contours were precise enough for recognition, they were not further refined. Fig. 8 shows two examples of eyebrow contours extracted. Because the shapes of eyes and mouth were not precise enough, we refined them using the PCDM approach stated in Section II-2. Fig. 9 shows two examples of eye and mouth contours extracted. After extracting the contours of eyebrows, eyes, and mouth in a face, we defined 30 FCPs to characterize the shape of the facial features and the normalization procedure of Sec. II-3 was then followed. Figs. 10 and 11, respectively, show the FCPs of the two examples of Figs. 9(a) and 9(b) before and after normalization. Then we chose six AUs, composed of FCP movements, as the input vectors for neural network-based expression classifiers.

Observing that a facial expression is mainly due to some local FCP movement, a radial basis function network (RBFN), which has local learning capability, could be very effective in recognizing a facial expression. Consequently, we first used RBFN to recognize the facial expressions. As shown in Fig. 12, the RBFN used for facial expression recognition is

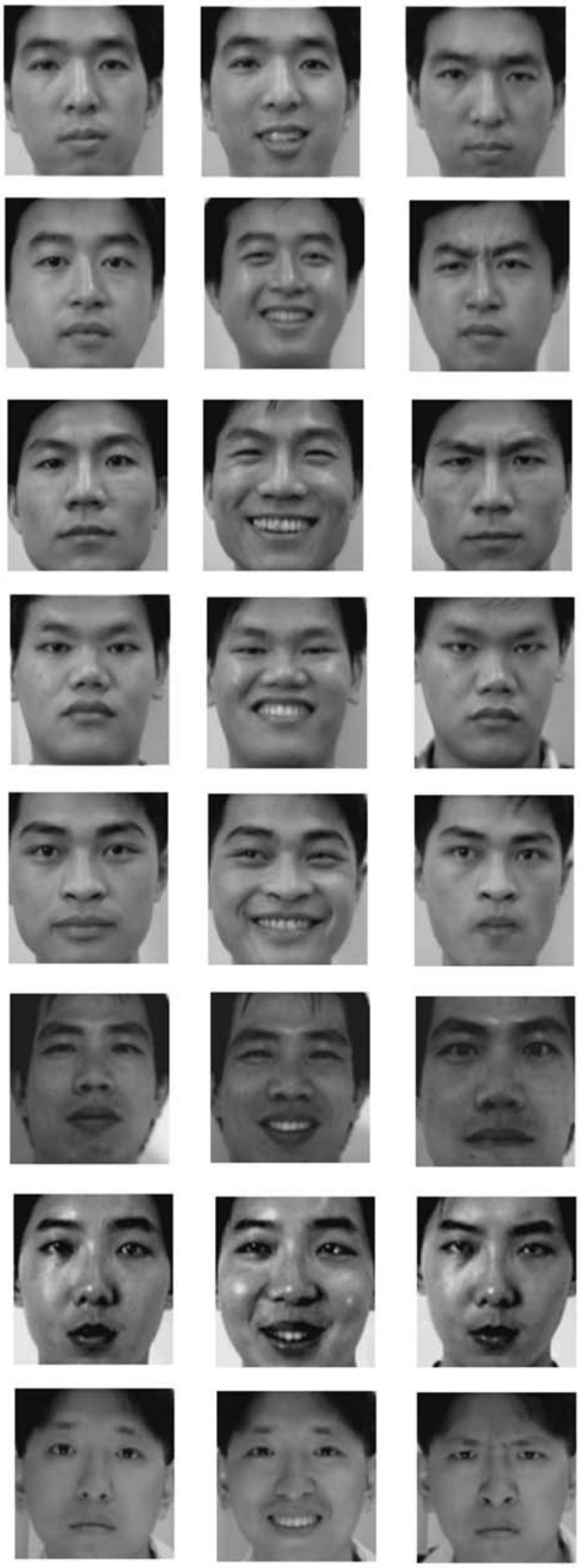
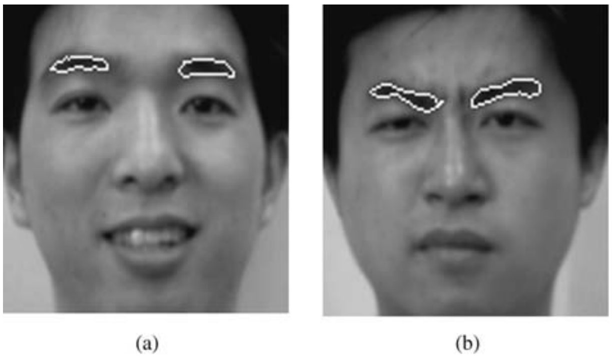
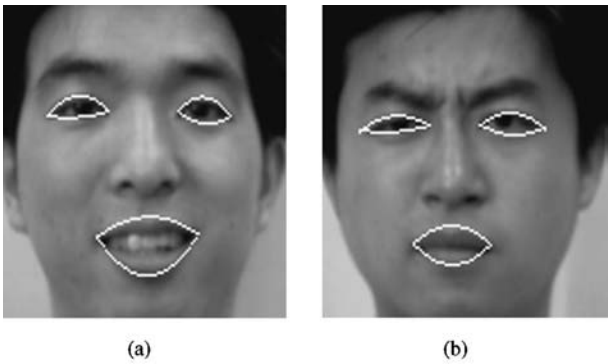


Fig. 7. Typical examples of neutral, happy, and angry images of eight persons in the database



(a) (b)

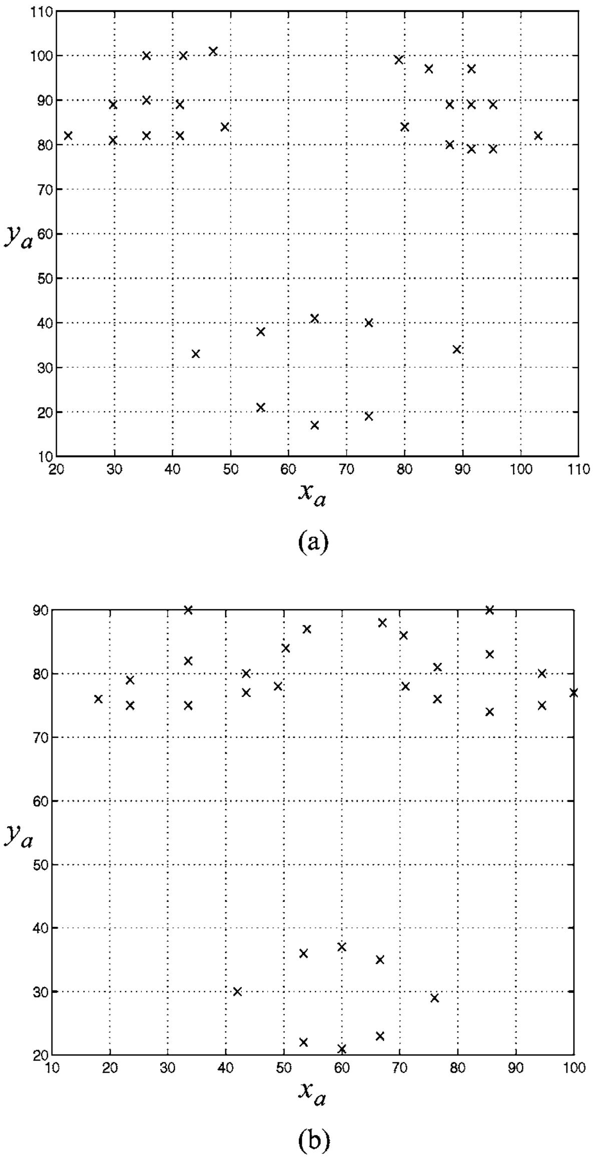
Fig. 8 Two examples of eyebrow contours extracted



(a) (b)

Fig. 9 Two examples of eye and mouth contours extracted

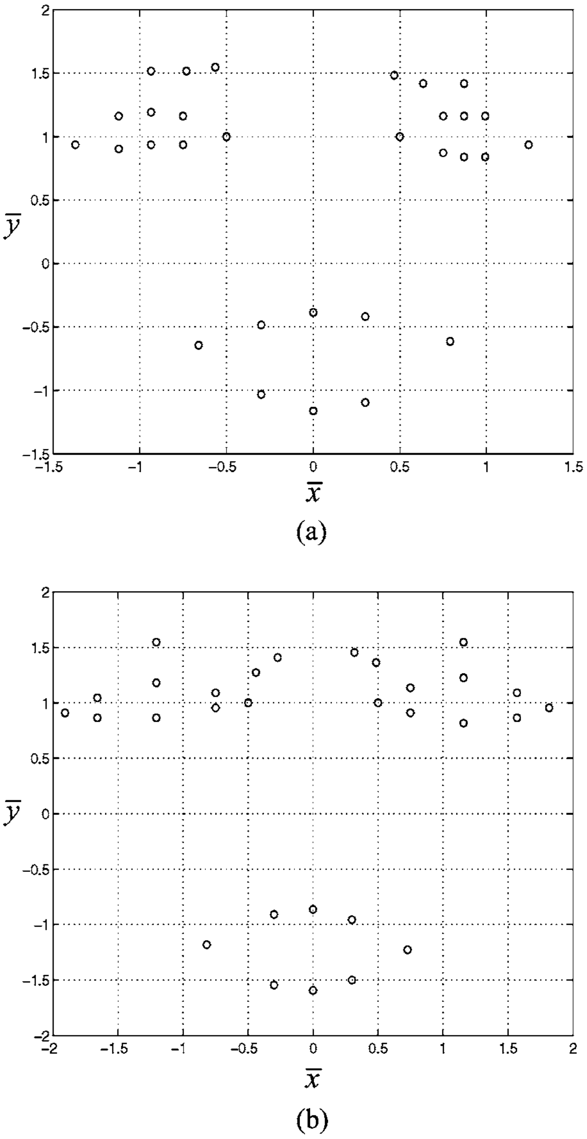
structured with six input neurons corresponding to six AUs, ten hidden nodes of Gaussian-shaped function, and three output neurons representing three facial expressions: neutral, anger, and happiness. The RBFN was trained by the back-propagation rule (Lin and Lee, 1996), where receptive field centers, widths, and connection weights are all learned supervisorily. After the training stage, 35 correct expressions were recognized among the 38 test images by the RBFN classifier, which leads to a recognition rate of 92.1%. For facial expression classification, we have also employed a multilayer perceptron (MLP) neural network trained by back-propagation learning rules (Chang *et al*., 1997). Embedded in an identical structure, a two-layer perceptron network with six inputs, ten hidden nodes, and three outputs was used. We have obtained the same recognition rate of 92. 1% for the MLP-based classifier. Although RBFN and MLP have reached the same recognition rate, the later usually converges more slowly and/or is apt to stick on local minima in the learning phase. It seems that a local-learning RBFN is more effective in recognizing a facial expression caused by local FCP movements. Compared with recognition of facial expressions from static images (Lanitis *et al*., 1997), we have obtained a better recognition rate by about 20%, disregarding the effect of classifying a smaller category of facial expressions. As noted before, recognizing a smaller, yet representative, expression category would be beneficial to obtain a more



(a)

(b)

Fig. 10 Two examples, (a) and (b), of the FCPs before normalization



(a)

(b)

Fig. 11 The normalized FCPs, (a) and (b), of Figs. 11(a) and 11 (b), respectively

reliable and consistent result with a minimal sacrifice to the need in practice.

V. CONCLUSION

In this paper, an automated facial expression recognition system using different neural network models, RBFN and MLP, has been described. We use the techniques of RCER and point contour detection method to obtain the features of the eyes, mouth, and eyebrows in a face. Thirty FCPs are defined to represent the position and thus the shape of the eyes, mouth, and eyebrows. AUs are then estimated by measuring the movements of FCPs. Finally, we recognize facial expressions by two neural network-based expression classifiers. We have obtained a high recognition rate of 92.1% by RBFN and MLP models. Although we have achieved the same recognition rate for these two neural models, the RBFN learns much faster than the back-propagation network mainly because a local-learning RBFN is well-suited for recognizing expression having local muscle movements. Simulation results by computers demonstrate that computers are capable of extracting high-level or abstract information like humans. The success of the abstract message of facial expression being recognizable by the computer demonstrates that other kinds of abstract information can still be recognized by a computer, which will result in much broader applications of the computer.

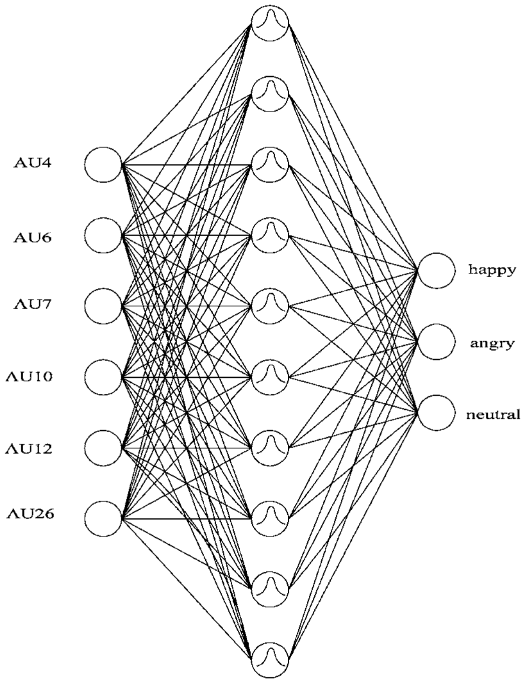


Fig. 12 The RBFN structure for facial expression recognition

ACKNOWLEDGMENT

This paper is supported in part by the National Science Council, Taiwan, Republic of China, under Grant NSC 87-2213-E-009-134. Gratitude also goes to the graduate students in the Department of Electrical and Control Engineering who contributed to the image data base of expressive "snapshots."

NOMENCLATURE

a_i	the i-th facial characteristic point.
base	the normalization parameter.
O_x, O_y	the origin of x''–y'' coordinate system.
x_a–y_b	the absolute coordinate system of FCPs.
x''–y''	the new coordinate system of FCPs.
(x_{ai}, y_{ai})	the coordinate of FCP a_i in x_a–y_a coordinate system.
(x''_{ai}, y''_{ai})	the coordinate of FCP a_i in x''–y'' coordinate system.
θ	the inclination of face with respect to the horizontal.
ϕ_{edge}	edge intensity image.

REFERENCES

1. Chang, J.Y., Chen, J.L., and Lin, J.F., 1997, "Towards an Automatic System for Facial Expression Recognition," *Proceedings, International Symposium on Multimedia Information Processing*, Taipei, Taiwan, R.O.C., pp. 129-134.
2. Chen, C.W., 1991, "Human Face Recognition Using Deformable Template and Active Contour," Master Thesis, National Tsing-Hua University, Hsin-Chu, Taiwan, R.O.C.
3. Cohn, J.F., Zlochower, A.J., Lien, J., and Kanade, T., 1999, "Automated Face Analysis by Feature Point Tracking Has High Concurrent Validity with Manual FACS Coding," *Psychophysiology*, Vol. 36, pp. 35-43.
4. Cootes, T.F., Taylor, C.J., Cooper, D.H., and Graham, J., 1995, "Active Shape Models - Their Training and Application," *Computer Vision and Image Understanding*, Vol. 61, No. 1, pp. 38-59.
5. Ekman, P., and Friesen, W.V., 1975, *Unmasking the Face*, Prentice-Hall, Englewood Cliffs, NJ.
6. Ekman, P., and Friesen, W.V., 1978, *The Facial Action Coding System*, Consulting Psychologist Press, San Francisco, CA.
7. Gonzalez, R.C., and Woods, R.E., 1992, *Digital Image Processing*, Addison Wesley, Reading, MA.
8. Howell, A.J., and Buxton, H., 1995, "Invariance in Radial Basis Function Neural Networks in Human Face Classification," *Neural Processing Letters*, Vol. 2, No. 3, pp. 26-30.
9. Huang, C.L., and Chen, C.W., 1992, "Human Facial Feature Extraction for Face Interpretation and Recognition," *Pattern Recognition*, Vol. 25, No. 12, pp. 1435-1444.
10. Kobayashi, H., and Hara, F., 1992, "Recognition of Six Basic Facial Expressions and Their Strength by Neural Network," *Proceedings, IEEE International Workshop on Robot and Human Communication*, New York, NY., pp. 381-386.
11. Kobayashi, H., and Hara, F., 1994, "Analysis of the neural network recognition characteristics of six basic facial expressions," *Proceedings, 3rd IEEE International Workshop on Robot and Human Communication*, New York, NY., pp. 222-227.
12. Lanitis, A., Taylor, C.J., and Cootes, T. F., 1997, "Automatic Interpretation and Coding of Face Images Using Flexible Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 743-756.
13. Lin, C.T., and Lee, C.S.G., 1996, *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice-Hall, Upper Saddle River, NJ.
14. Mase, K., 1991, "Recognition of Facial Expression From Optical flow," *IEICE Transactions*, Vol. 74, No. 10, pp. 3474-3483.
15. Matsuno, K., Lee, C.W., Kimura, S., and Tsuji, S., 1995, "Automatic Recognition of Human Facial Expressions," *Proceeding, Fifth International Conference on Computer Vision*, Los Alamitos, CA., pp. 352-359.

16. Morishima, S., and Harashima, H., 1993, “Emotion Space for Analysis and Synthesis of Facial Expression,” *Proceedings, 2nd IEEE International Workshop on Robot and Human Communication*, New York, NY., pp. 188-193.
17. Padgett, C., and Cottrell, G., 1997, “Representing Face Images for Emotion Classification,” *Advances in Neural Information Processing System*, Vol. 9, pp. 894-900.
18. Rosenblum, M., Yacoob, Y., and Davis, L.S., 1996, “Human Expression Recognition from Motion Using a Radial Basis Function Network Architecture,” *IEEE Transactions on Neural Networks*, Vol. 7, No. 5, pp. 1121-1138.
19. Shackleton, M.A., and Welsh, W.J., 1991, “Classification of Facial Features for Recognition,” *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Los Alamitos, CA., pp. 573-579.
20. Sirovich, L., and Kirby, M., 1987, “Low-Dimensional Procedure for the Characterization of Human Faces,” *Journal of the Optical Society of America Part A -Optical and Image Science*, Vol. 4, No. 3, pp. 519-524.
21. Turk, M.A., and Pentland, A.P., 1991a, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86.
22. Turk, M.A., and Pentland, A.P., 1991b, “Face Recognition Using Eigenfaces,” *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Los Alamitos, CA., pp. 586-591.
23. Yuille, A.L., Cohen, D.S., and Hallinan, P.W., 1989, “Feature Extraction from Faces Using Deformable Templates,” *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC., pp. 104-109.
24. Yuille, A.L., Hallinan, P.W., and Cohen, D.S., 1992, “Feature Extraction from Faces Using Deformable Templates,” *International Journal of Computer Vision*, Vol. 8, No. 2, pp. 99-111.
25. Zhang, Z., 1999, “Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiments with a Multilayer Perceptron,” *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 13, No. 6, pp. 893-911.

Manuscript Received: Apr. 12, 2000
Revision Received: Sep. 05, 2000
and Accepted: Oct. 29, 2000

利用類神經網路於臉部表情自動辨識系統

張志永　陳嘉麟

國立交通大學電機與控制工程學系

摘　要

本論文提出利用類神經分類器之臉部表情自動辨識系統。首先，我們使用粗略輪廓預測程式（rough contour estimation routine）、數學形態學（mathematical morphology）、以及點輪廓偵測法（point contour detection method)等影像處理的技術，來擷取眉毛、眼睛、及嘴巴這三個特徵器官的正確輪廓。接著再定義30個臉部特徵點(facial characteristic points)來描述上述三個特徵器官的位置和形狀，並產生運動單元(action units)來代表人臉基本的肌肉運動，因此臉部表情可以藉由這些運動單元的組合來表示。我們選取六個主要的運動單元當做以類神經網路爲基礎的表情分類器之輸入向量，而這六個運動單元是由臉部特徵點的變化所組合而成。本論文所採用的表情分類器分別爲放射狀基礎函數網路（radial basis function network）以及多層認知網路（multi-layer perceptron network）。使用上述兩種類神經網路來辨識平常、生氣、以及高興等三種表情皆可獲得高達92.1%的辨識率。經由此電腦模擬的結果顯示電腦可以像人類一樣擷取高階或抽象的資訊。

關鍵詞：臉部表情辨識，類神經分類器，點輪廓偵測法，臉部運動單元。