Facial Expression Synthesis by Radial Basis Function Network and Image Warping *

Mary Y.Y. Leung, Hung Yen Hui and Irwin King
{yyleung, yhhung, king}@cs.cuhk.hk
Department of CS & E, The Chinese University of Hong Kong

ABSTRACT

This paper presents a novel approach in synthesizing various facial expressions by image warping, using spatial displacement of facial landmark point generated by Radial Basis Function (RBF) networks. We trained two RBF networks from test images to obtain these spatial displacements. One RBF network can be used to generate different degrees of expressions, while the other can be used to obtain mixed facial expressions. We discuss the method used and demonstrate the results.

1. Introduction

Facial expressions play an important role in non-verbal communications. Many applications for teleconferencing, human computer interface and computer animation require realistic reproduction of facial expressions.

Suppose we are given a normal face image of a person, i.e., a face image without any expression of emotion, how can we synthesize different expressions of that person? Current approaches in synthesizing facial expressions include the use of 3D model of facial muscles and tissues [4], texture mapping approach to 3D facial image synthesis [10]. This method proves to be tedious in determining the actual parameter values for synthesizing and animating facial expressions. An alternative approach has been investigated by Nur Arad et al. [8], which demonstrated the use of radial basis function in interpolating the anchor points for 2D image warping, which can be applied to synthesize facial expressions. However, it does not provide a mechanism to determine the appropriate destination of the anchor points for each particular facial expression.

This paper deals with 2D image warping in synthesizing human facial expressions. We have made use of 30 landmarks obtained from the face, which have greater potentials in revealing changes in displaying a particular facial expression. The differences in movements of these landmarks for the facial expressions are used as control points in image warping for synthesizing various facial expressions. We construct the neural network description of the mapping relationship between necessary patterns of movements of these landmarks and the 6 universal facial expressions [3]. The distinctive difference between this method and [8] is that, we use the RBF in the network description for finding the necessary changes of the landmarks, rather than in interpolating the anchor points in image warping.

2. Reverse of Facial Expression Recognition

The synthesis of facial expressions can be seen as a reverse process of facial expressions recognition. In recognition, we present the necessary information (the movements of the landmarks) so as to classify for a particular facial expression (an emotion label). But what if we reverse the question: given a particular facial expression, can you tell what the necessary movements of the landmarks are? This indeed can be seen as a reverse process to which emotion labels are used as inputs and the outputs are the movements of the landmarks Figure 1(a).

*The research described in this paper was supported in part by a RGC Grant 221500620, Direct Grant 220500910 and Direct Grant 220500720.

0-7803-3210-5/96 $4.00©1996 IEEE

2.1. Recognition

In [7], the method of recognizing the 6 universal facial expressions using neural network is discussed. A brief summary is presented as follows.

A set of 30 landmarks located near the eye-brows, eyes and the mouth are defined as the Facial Characteristic Points(FCPs). These points are obtained manually by mouse clicks and are shown in Figure 1(b). They are selected as they have greater potential in revealing the changes in showing facial expressions.

The basic idea is to find out the differences between the FCPs of the normal face and that of the expressive face. Thus, the differences of those 30 pairs of position information will constitute the 60 inputs to the neural network. The number of output layer unit is 6, the position of which corresponds to each of the 6 emotion labels in the order `happy`, `sad`, `angry`, `fear`, `surprised` and `disgusted`.

To compensate for the differences in the size, orientation and position of the faces in the image as well as the size of the face components, we have to process the coordinates of the FCPs so that they are comparable. Therefore, the absolute coordinates of the FCPs have to undergo 4 steps:

Translation - It is employed to translate the origin of coordinate system to the nose top of the individual as the absolute pixel coordinatess of the FCPs are obtained relative to the lower left corner of the image. A quantity called *base* is introduced, which seems not to vary for each of the facial expressions,

$$base = \sqrt{(xb_2 - xb_1)^2 + (yb_2 - yb1)^2} \tag{1}$$

where (xb_1, yb_1), (xb_2, yb_2) are the pixel position of inner corners of left eye and right eye respectively.

The inclination of the face with respect to the horizontal line, θ,is introduced. The mid-point (x_0, y_0) between (xb_1, yb_1) and (xb_2, yb_2) is also found using the mid-point formula. The origin of the new coordinate system $(origin_x, origin_y)$ is calculated as $(x_0 - base * \sin\theta, y_0 - base * \cos\theta)$.

Rotation - It is employed to correct the inclination of the face so that the coordinates are expressed with respect to the vertical axis of the face in the new coordinate system.

Normalization - It is introduced to compensate the distance effect between the client's face and the camera. Both of the landmarks of the normal and expressive face after rotation are divided by the value *base*. These normalized values of the expressive face are subtracted from those of the normal one.

Standardization - As those subtracted values are indeed the absolute displacements of the FCPs from their normal position, thus these magnitudes are subject to individual variations. Standardization is needed to find out the relative displacement of the landmarks from their normal position.

2.2. Synthesis as a reverse process

2.2.1. The Radial Basis Function (RBF) Network

The basic principle of synthesizing facial expressions is to find out the necessary relative spatial shift of the FCPs for each expression of emotion. So what is initially the input to the neural network in recognition will become the output in the synthesis and vice versa. We employed a RBF network to solve this problem.

Radial Basis Functions have proven to be an effective tool in interpolating data in multidimensional spaces. Given a set of scattered n-dimensional data, $(\vec{x}_i, \vec{f}_i), \vec{x}_i \in R^n, \vec{f}_i \in R^{n'}$ for $i = 1, ..., m$, in order to choose a function $s : R^n \mapsto R^{n'}$ which satisfies the interpolation conditions

$$s_k(\vec{x}_i) = f_{ik} \quad i = 1, ..., m \quad k = 1, ..., n' \tag{2}$$

we would consider interpolating functions of the form

$$s_k(\vec{x}) = \sum_{j=1}^{m} w_{jk} g(\| \vec{x} - \vec{\mu}_j \|), \quad \vec{x} \in R^n, \quad k = 1, 2, ..., n' \tag{3}$$

where $\| \bullet \|$ denotes the usual Euclidean norm on R^n.

In particular, the $g()$ we employed is the normalized Gaussian activation function:

$$g(\vec{x}) = \frac{\exp[-(\vec{x} - \vec{\mu}_j)^2 / 2\sigma_j^2]}{\sum_k \exp[-(\vec{x} - \vec{\mu}_k)^2 / 2\sigma_k^2]} \tag{4}$$

where x is the input vector, μ is a set of weights and σ is the width of the RBF.

Thus, the determination of the nonlinear map $s(\vec{x})$ has been reduced to the problem of solving the following set of linear equations for the coefficients w_j,

$$\begin{pmatrix} f_{1k} \\ \vdots \\ f_{mk} \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mm} \end{pmatrix} \begin{pmatrix} w_{1k} \\ \vdots \\ w_{mk} \end{pmatrix}, k = 1, 2, ..., n'$$

where $A_{ij} = g(\| \vec{x}_i - \vec{\mu}_j \|)\ i, j = 1, 2, ..., m$.

The architecture of our RBF network is shown in Figure 1(c). Our intention is to find out a mapping between the emotion label and the movements of the FCPs. Training the RBF network is equivalent to solving for the mapping that interpolates the training data. The x in the Gaussian function corresponds to the emotion label of the neural network while the output of the network is a vector of movements of FCPs. The σ corresponds to the spread constants set in the training setup. The only real design decision for RBF network is to find a good value of σ, which determines the generalization ability of the RBFs. The variable σ should be large enough to allow the overlapping of the input regions of radial basis functions. This makes the network function smoother and results in better generalization for new input vectors occurring between input vectors. However, σ should not be so large that each neuron responds in essentially the same manner, i.e., any information presented to the network becomes lost. Our σ is picked by trial and error, within maximum and minimum of distances of input vectors. After training, 2 sets of weights, μ and w, will be obtained. This is used to produce the spatial displacement.

2.2.2. Mapping the Output to the Image

As the output from the RBF network is a vector that consists of relative displacements of FCPs representing certain facial expression, they have to be de-standardized, de-normalized and transformed back according to the FCPs of the normal face, so that we can obtain their locations on the normal face image.

To de-standardize, the elements of the output vector are multiplied by their corresponding standard values as defined in [7] and then added to the normalized values of the FCPs of the normal face image:

$$x_i := x_i + normal\_x_i, \qquad y_i := y_i + normal\_y_i \tag{5}$$

These values are then de-normalized by the *base* value:

$$rotx_i := x_i * base, \qquad roty_i := y_i * base \tag{6}$$

and then rotated and translated back:

$$x_i := rotx_i * \cos\theta - roty_i * \sin\theta, \qquad y_i := rotx_i * \sin\theta + roty_i * \cos\theta \tag{7}$$

$$xb_i := x_i + origin_x, \qquad yb_i := y_i + origin_y \tag{8}$$

(xb_i, yb_i) is then the new position of the ith FCP on the normal face image.

To generate an expressive face image from the normal face image, we employed 2D image warping [5]. Both of the FCPs of the normal face and that of the expressive face are connected to form triangular patches as shown on Figure 1(d). Image warping is then performed by scan-converting each triangle.

2.3. Experimental Results

We have carried out 2 sets of training using the Neural Network Toolbox of Matlab running on Sun Sparc20. A set of 128×128 grayscale images are used in our experiment. The details of the system setup are summarized in Table 1.

	σ	no. of images	CPU time to train (sec.)
RBN1	0.5	30	3.60
RBN2	1	6	0.79

Table 1: Summary of the setup of the 2 trainings

2.3.1. Training with different degrees of 6 universal facial expressions

Facial expressions can have different strengths, e.g, the degree of happiness ranges from smile to grin to laugh. Therefore, we used five images with different degrees for each of the 6 facial expressions, thus a total of 30 images are used as the training set. For each expression, we manually arrange the images in the order of decreasing strength and assign to each of them the value 1.0, 0.9, 0.8, 0.7 and 0.6 accordingly. Thus the input vector would be in the form (0.6,0,0,0,0,0) for the weakest degree of happiness among the 5 images.

After training, the network is able to generate the 6 universal facial expressions, plus different degrees of variation for each facial expression (RBN1). From Figure 2(a), we can see that the neural network is able to capture the features of the facial expressions: for happy face, the upward movement of mouth corners is captured; for sad face, the inner eyebrows raise and mouth corners go down; for surprise the whole eyebrows raise and the mouth widely open, etc. Figure 2(b) shows 5 degrees of happy faces generated. We find that this process is not a linear one.

2.3.2. Training with a set of 6 universal facial expressions

We find that the above network is unable to generate mixed expressions, i.e., if we specify the input as a mix of some degrees of sadness and happiness, the network cannot generate the output as a mixed expression. Rather, it constantly gives a particular set of 60-element output for all mixed inputs. We train another RBF network with a set of 6 universal expressions (RBN2). When mixed inputs such as $(1.0, 1.0, 0, 0, 0, 0)$ are applied, the output shows a dependence on the mixed input, resulting in a mixed expression. However, this network is unable to generate different degree for each expression. Figure 2(c) and Figure 2(d) show the results obtained from this network.

3. Discussions

From our experimental results, the main defect is that we cannot find a single RBF network that is able to generate both mixed expressions and different degree for each expression.

We have mentioned that the input vectors of RBN1 consist of arbitrarily assigned values that represent the degree of emotion. These distinct data points are required for different degrees of variation for each expression. Therefore there is a need to adjust the inputs for the training to be successful. This is a highly subjective judgement.

Apart from the movements and shapes of the facial components, wrinkles such as naso-labial folds or crow's-feet wrinkles on the face contribute significantly to facial expressions. However, our system only uses normal face images where the face is free from any wrinkle, and simple warping technique is employed. Special technique such as texture mapping should be adopted to add such wrinkles to the final image. Also, new FCPs need to be added on eyebrows and mouth so as to obtain smooth warped eyebrows, and an open mouth can be modelled from a normal face image with close mouth. In the future, we would like to automate the process of obtaining the landmarks from face images.

4. Conclusion

This paper aims at building a system for synthesizing facial expression of emotions. Radial Basis Function networks are used to map the emotion labels to the displacements of the 30 FCPs. Depending on the

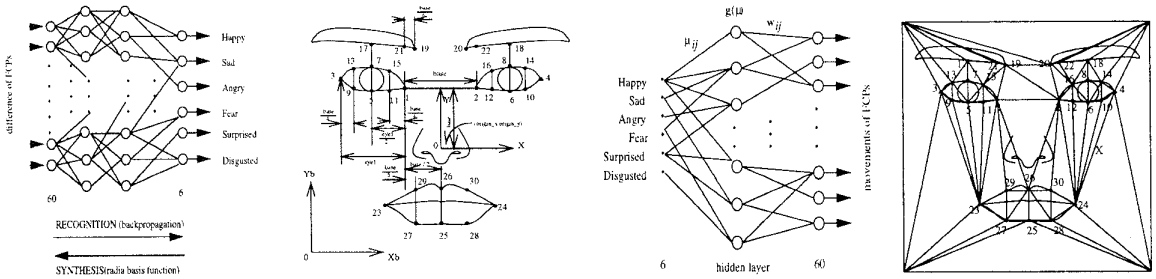

Figure 1: (a) The neural networks for recognition and synthesis have reverse input and output. (b) The 30 facial characteristic points. (c) The architecture of RBF network. (d) The FCPs are connected to form patches.

positions of the sample data points, the RBF approach constructs a linear function space according to an arbitrary distance measure. Thus an interpolating surface which exactly passes through all the pairs of the training set can be produced so that when data points not in the training set are presented to the RBF network, the mapping can also be interpolated. Since facial expressions have different degrees and are often mixed with one another, the use of RBF network for interpolation makes generation of these kinds of facial expressions possible.

Although 3D model is a way of avoiding distortion due to 2D image warping, we still employ a 2D model as it is applicable to different individuals. In order to achieve a more natural look of the synthesized facial expression from a normal face image, additional texture mapping can be used to introduce facial texture such as wrinkles on bridge of nose and the naso-labial folds. Lastly, we may need additional FCPs for synthesizing more realistic facial images.

References

[1] P. J. Benson. Morph transformation of the face image. *Image and Vision Computing*, 12:691–696, 1994.

[2] D. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

[3] P. Ekman and W. V. Friesen. *Unmasking the Face.* Consulting Psychologists Press, Inc., 1975.

[4] F. Hara and H. Kobayashi. Computer graphics for expressing robot-artificial emotions. IEEE International workshop on Robot and Human Communication, 1992.

[5] P. Heckbert. Graphics gems. pages 65–77, 1990.

[6] K. Hertz and Palmer. *Introduction to the theory of neural computation.* Addison Wesley, 1991.

[7] H. Kobayashi and F. Hara. Recognition of six basic facial expressions and their strength by neural network. IEEE International workshop on Robot and Human Communication, 1992.

[8] D. R. Nur Arad, Nira Dyn and Y. Yeshurun. Image warping by radial basis functions: Application to facial expressions. *CVGIP: Graphical Models and Image Processing*, 56, No. 2:161–172, 1994.

[9] G. Wolberg. *Digital Image Warping.* IEEE Computer Society Press Monograph, 1990.

[10] J. Yau and A. Duffy. *A Texture mapping approach to 3D facial image synthesis.* 1988.

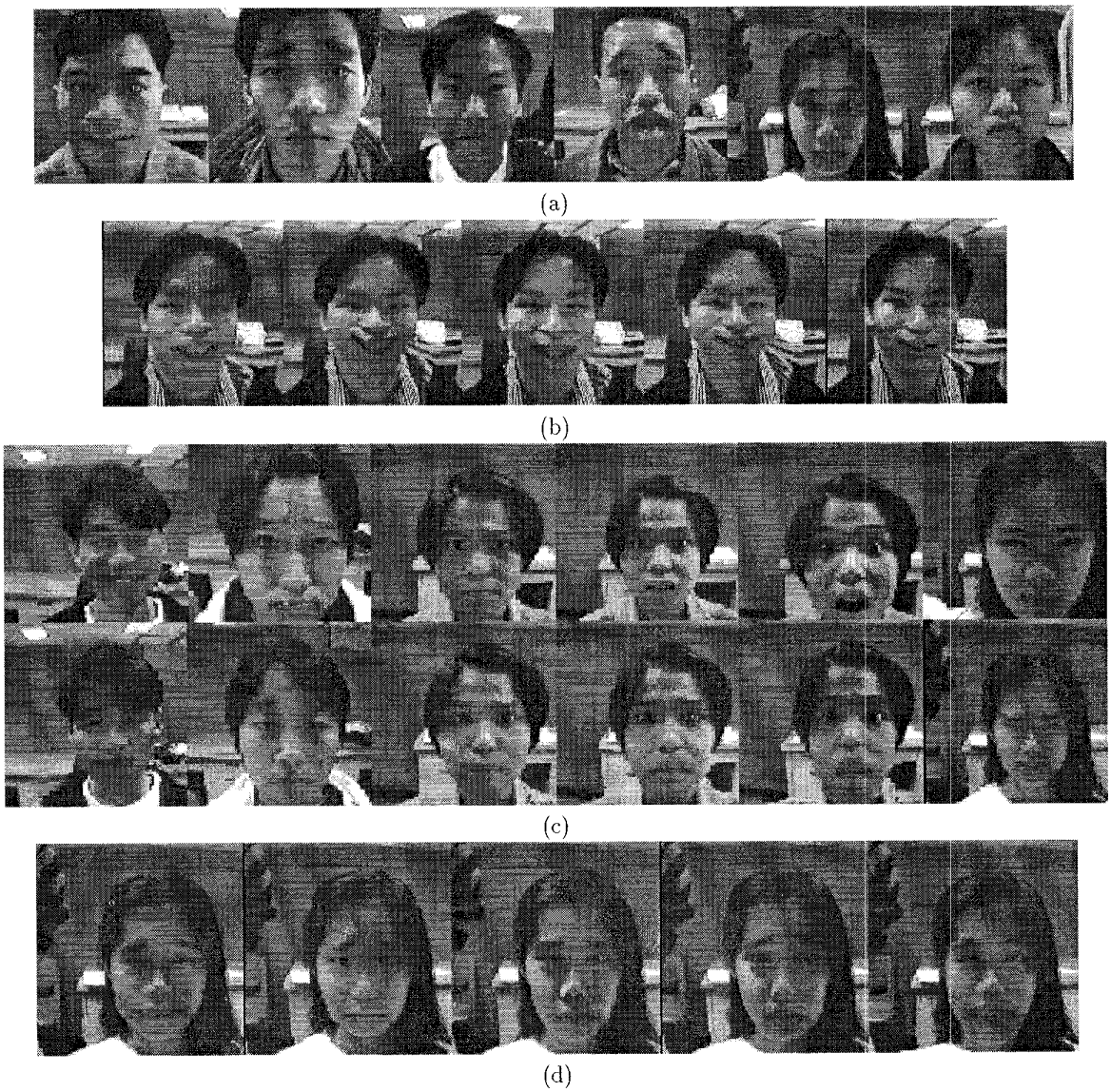

Figure 2: (a) The 6 universal facial expressions generated by RBN1. They are warped images showing, from left to right, the expression of `happy`, `sad`, `angry`, `fear`, `surprised` and `disgusted` respectively. (b) 5 degrees of happy faces generated by RBN1. The degree of happiness is decreasing from left to right. (c) The 6 universal facial expressions generated by RBN2. The 1st row shows how the clients respond to the request to exhibit the 6 universal expressions, the 2nd row shows the images warped with the FCPs obtained by using RBN2. (d) mixed expressions generated by RBN2. From left to right, the image shows `happy` mixed with `sad`, `sad` mixed with `angry`, `angry` mixed with `fear`, `fear` mixed with `surprised` and `happy` mixed with `surprised`.