MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

3—29 Partial Face Extraction and Recognition Using Radial Basis Function Networks

Nan He* Kiminori Sato* Yukitoshi Takahashi*
Department of Information and Computer Science
Kagoshima University

Abstract

Partial face images, e.g., eyes, nose, and ear images are significant for face recognition. In this paper, we present a method for partial face extraction and recognition based on Radial Basis Function (RBF) networks. Focus has been centered on using ear images because they are not influenced by facial expression, and the influences of aging are negligible. Original human side face image with 320×240 pixels is input, and then the RBF network locates the ear and extracts it with a 200×120 pixels image. Next, another RBF network is constructed for the purpose of recognition. An algorithm that determines the radius of an RBF function is introduced. Dynamic radius, so called as compared to static one, is found through the algorithm that makes RBF functions adaptable to the training samples. We build a database that contains 100 people's 6 side face images to test the method and the results of both extraction and recognition are satisfied.

1 RBF networks with Dynamic Radius

Radial basis function networks can be regarded as a multivariate interpolation problem [1]. Therefore, learning is equivalent to finding a surface in a multi-dimensional space that provides a best fit to the training sample, and generalization is equivalent to the use of this multi-demensional surface to interpolate the test data. In 1990, Poggio and Girosi applied the regularization theory[2] to this class of networks as a method for improved generalization to new data[3].

As depicted in Figure 1, the construction of a RBF network involves three layers with entirely different roles. The input layer is consisted of source nodes that connect the network to its environment. The second layer, the only hidden layer in the network, applies a nonlinear transformation from the input space to the hidden space. The output layer supplies the response of the network to the activation pattern.

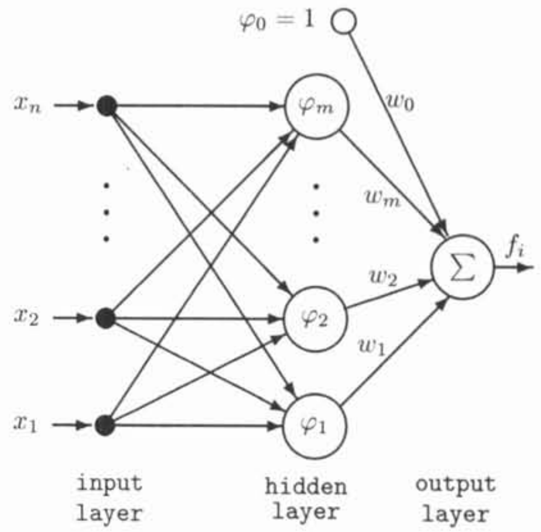


Figure 1. Radial Basis Function Networks

When a RBF network is used to approximate function $f : \Re^n \to \Re^p$, several parameters should be appropriately adjusted in the formula

$$f_i(\mathbf{x}) = \sum_{j=0}^{m} w_{ij}\varphi_j(\mathbf{x}) \qquad (i = 1, \cdots, p)$$

where $\mathbf{x} = [x_1, \cdots, x_n]^T$ is an input vector, and

$$\varphi_j(\mathbf{x}) = exp(-\frac{\|\mathbf{x} - \mathbf{c}_j\|^2}{\mathbf{r}_j^2}) \quad (j = 1, \cdots, m)$$

is a RBF function, commonly a Gaussian, that centered at $\mathbf{c}_j = [c_{1j}, \cdots, c_{nj}]^{\mathbf{T}}$ and has the radius $\mathbf{r}_j = [r_{1j}, \cdots, r_{nj}]^{\mathbf{T}}$. We define

$$r_{lj} = \frac{1}{\sqrt{2m}} \sum_{k \neq j} |c_{lk} - c_{lj}| \qquad (1)$$

*Address: 1-21-40 Korimoto, Kagoshima-shi, Kagoshima 890-0065 Japan. E-mail: (ganan, kimi, takahasi)@ics.kagoshima-u.ac.jp

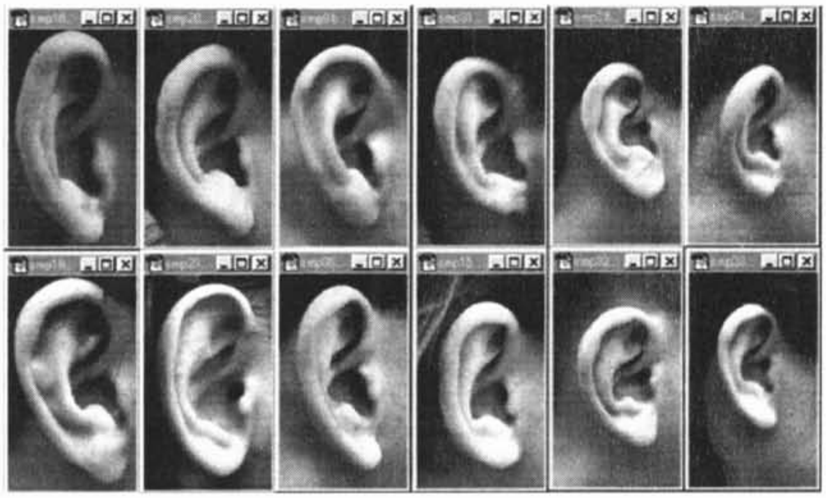


Figure 3. Training Set for Extraction

as Dynamic Radius, where k and $j = 1, \cdots, m$, $l = 1, \cdots, n$. The meaning of Dynamic Radius is that the radius of each RBF function is determined not only by its own position but also by the positions of other RBF functions.

Figure 2. Dynamic Radius

The properties of dynamic radius were revealed in more detail in Figure 2 in which three RBF functions are displayed in a two-dimension space. In each image there are two RBF functions on the left side, and one RBF function on the right side. When the right-hand RBF Function is moved away from the left-hand two, shown in the second and third images, the radii of all the three Functions are changed. The alteration of radius also means the alteration of reception field of RBF function. And Dynamic Radii try to segregate all the RBF functions through the adjustment of their reception fields.

To build a RBF network, a RBF function will be centered at every sample of the training set, and assigned its radius calculated by Equation (1). Also, we need to determine the output layer weights, w_{ij}. Suppose the transformed training samples have been arranged in a matrix, $\mathbf{H}$, called design matrix, so that each row represents one training sample and the j-th column contains the value of φ_j for the sample. One more column is added standing for φ_0, whose value is always 1.

$$\mathbf{H} = \begin{bmatrix} \varphi_0 & \varphi_1(\mathbf{x}_1) & \cdots & \varphi_m(\mathbf{x}_1) \\ \varphi_0 & \varphi_1(\mathbf{x}_2) & \cdots & \varphi_m(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_0 & \varphi_1(\mathbf{x}_p) & \cdots & \varphi_m(\mathbf{x}_p) \end{bmatrix}^{p\times(m+1)}$$

Let $\mathbf{d}$ denote the classification matrix, sometimes called the teacher signals, where each column represents one concept label. Having $\mathbf{H}$ and $\mathbf{d}$, the weight matrix, $\mathbf{w}$, can be found by using the well-known equation

$$\mathbf{w} = (\mathbf{H}^{\mathbf{T}}\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}\mathbf{d}.$$

Finally, we define the network's output as

$$y = \frac{f_{max}}{\|\mathbf{F}\|} \tag{2}$$

where f_{max} is the maximum element of vector $\mathbf{F} = [f_1, \cdots, f_p]^{\mathbf{T}}$. The output y, defined in Equation (2), can be regarded as an angular measurement in the output p-dimension space. The threshold set for y is usually 0.6.

2 Extraction

Image data were obtained under uniform lighting conditions. The distance between a subject and the video camera was fixed. Six images per subject were acquired. In order to get natural variations, the subject (ear image) was separated from camera after each acquisition. In the case of extraction, the training set of the RBF network is required to be sensible to different kinds of side face images that are available in our database. We selected the training samples that have 200×120 pixels through experiments. Figure 3 shows the training set in which ears with different sizes were included. To speed up our programs, a mosaic process is applied to reduce the image size from 200×120 pixels to 20×12 pixels. So the training set can be expressed as $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_j, \cdots, \mathbf{x}_{12}]^{\mathbf{T}}$ where each $\mathbf{x}_j$ possesses 240 elements, also means the input space has 240 dimensions. Constructed through the method depicted in section 1, the RBF network contains 12 RBF functions, each of them has 240 source nodes. And the centers of these RBF functions are $\mathbf{c}_j = \mathbf{x}_j$, and their radii

$$\mathbf{R} = [\mathbf{r}_1, \cdots, \mathbf{r}_j, \cdots, \mathbf{r}_{12}]^{\mathbf{T}},$$

where $j = 1, 2, \cdots, 12$ and each $\mathbf{r}_j$ is determined by Equation (1).

Corresponding to the training set shown in Figure 3, the classification matrix $\mathbf{d}$, shown below, assigns the ears that have almost the same size with the same concept label, and also it define the output space's 6 dimensions.

$$\mathbf{d} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{bmatrix}^{12\times 6}$$

The extraction is a reiterative processing, like the SSDA method[4]. In one reiteration, a part of the human side face image, with 200×120 pixels, is collected as an input data. The RBF network return an output y defined in Equation (2). Arranging all the outputs according to the position of their input samples in the side face image, we get a surface, called mapping surface. Or we may say that the RBF network maps a side face image to a surface in which the most steep peak implies the location of the ear in the side face image.

It is notable that sometimes the point of the steep peak doesn't hold the maximum value; it is hidden in the mapping surface, and cannot be able to find out using a simple method. To handle this problem, we use an edge detecting technique in the field of image processing. Convoluting the mapping surface with Laplacian Mask, the mapping surface becomes plainer and the steep peak always rises.

0	-1	0
-1	4	-1
0	-1	0

Laplacian Mask

Figure 4 illustrates the procedure that processing a side face image to a mapping surface through which we can extract the ear image that has 200×120 pixels.

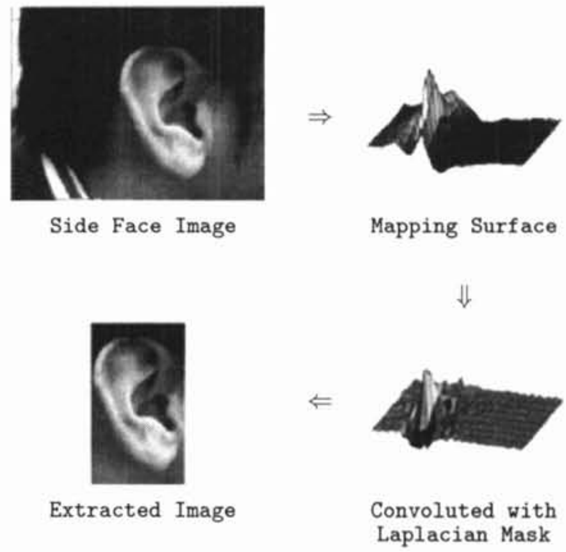


Figure 4. Process of Extraction

For the extracted images, there are unavoidable location differences between the ears that belong to same person. Fortunately, benefited from the dynamic radius algorithm, these differences were kept within a small range to ensure the success of future recognition.

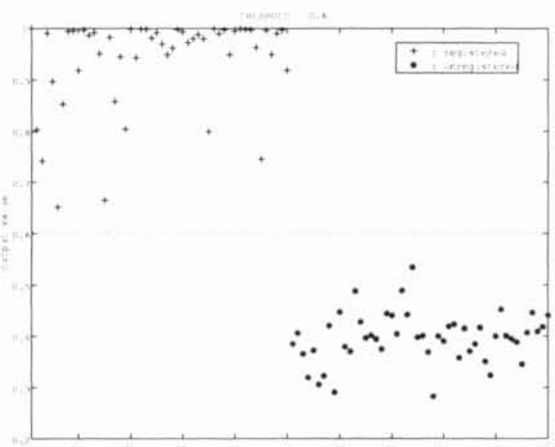
3 Recognition

The extraction algorithm automatically generated a new database which contains 600 ear images, from 100 people. The data size of each partial face images was reduced through the mosaic process. The recognition begins with the construction of the training set for its RBF network. Those people whose ear images were included in the training set were called registered people. Usually, for each registered person, three of his/her ear images were used to train the network, and all the others were used for test. The network's output value is defined in Equation (2).

Several experiments were conducted to test our method. Figure 5 and Figure 6 show the output pictures of experiments in which 50 and 70 people were registered respectively. From these figures, we can find that all the output values of registered people are larger than the threshold, and the values of unregistered people are smaller than the threshold. In both results, between output values of the two different groups, enough ranges have been obtained to recognize whether each person is registered or not. Then we get a 100% recognition rate for both registered and unregistered people[5]. To check the efficiency of dynamic radius, we use the static radius to do the same experiments: extraction and recognition. Figure 7 shows the output picture in which there are several unsuccessful points. The reason is that the database, generated by extraction stage using static radius, contains some failure extracted ear-images.

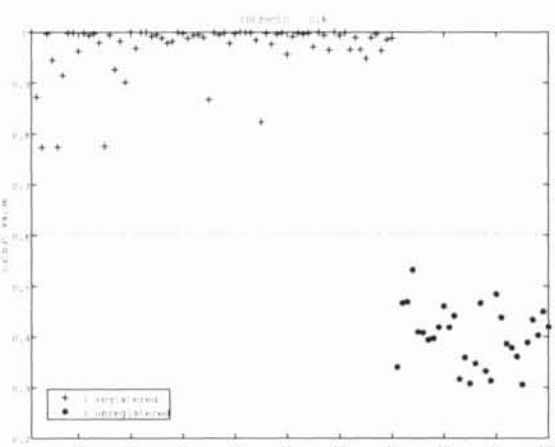
4 Conclusion

We successfully extracted and recognized the partial face images by using RBF network with dynamic radius algorithm proposed in this paper. The recognition rates for both registered and unregistered people are 100%.



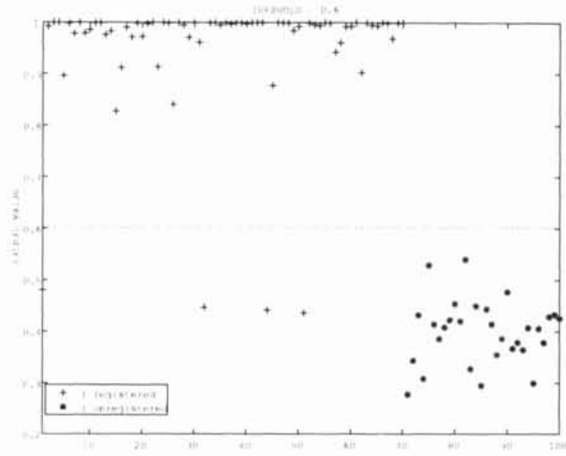
`Registered 50 people, dynamic radius`

Figure 5. Experiment 1



`Registered 70 people, dynamic radius`

Figure 6. Experiment 2



`Registered 70 people, static radius`

Figure 7. Experiment 3

References

[1] Powell, M.J.D., "Radial basis functions for multivariable interpolation: A review," *IMA Conference on Algorithms for the Approximation of Function and Data*, pp. 143-167, RMCS, Shrivenham, England. 1985.

[2] Tikhonov, A.N., "On solving incorrectly posed problems and method of regularization," *Doklady Akademii Nauk USSR*, vol. 151, pp. 501-504. 1963.

[3] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proceedings of the IEEE* vol. 78, pp. 1481-1497. 1990.

[4] M. Sone, M. Sakauchi and M. Onoe, "A note on image registration based on geometric correspondence between two levels of extended pyramid structure" *IEICE trans. D*, vol. J71-D, No.7, pp.1238-1248, 1988.

[5] K. Sato, S. Shah and J. K. Aggarwal "Partial face recognition using Radial Basis Function Networks". *International conference on automatic face and gesture recognition*, FG98, pp.288-293, 1998.