ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 395-400

Grouping Complex Face Parts by Nonlinear Oscillations

Souichi Oka, Miki Kitabata, Yoshiaki Ajioka and Yoshiyasu Takefuji

Graduate School of Media and Governance, Keio University,
5322 Endo, Fujisawa, 252, JAPAN
email: souichi@sfc.keio.ac.jp

Abstract.
This paper proposes a new neural network model for human face detection. After face parts such as eyes and mouths are extracted from target images in a pre-processing system, they are fed to a recurrent neural network composed of nonlinear oscillations system. The proposed neural network can detect several faces one after another by composing face parts extracted by pre-processing. A whole system works well even when some face parts are lost.

1. Introduction

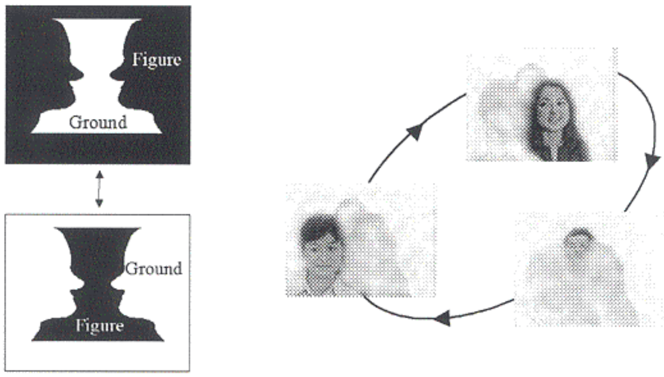
One of the most difficult problems in visual perception is the *figure-ground* separation. Our brains can separate it into figure-segments regarded as an object and ground-segments as a background immediately. This function enables us to recognize two objects in a scene(see Fig. 1(a)).

Gestalt psychology suggests that we must recognize an object by grouping local features detected by the primary visual area. Since the grouping must be performed in recognizing each of objects if so, it is considered that figure-ground separation is caused by the grouping.

In the 1980's, some neural physiologists have suggested that figure-ground separation is caused by intermittent excitatory and inhibitory behavior of neurons in our brain[1][2]. It is derived that such behavior of neurons realizes grouping. Although some researchers have proposed some interesting models of neural networks in order to recognize images, their models do not work well because they do not concern with normalization on size, rotation and shift of an object by grouping its features[1]. Our system realizes shch a normalization by grouping, bacause grouping is just to combine face parts simply.

A goal of this paper is to recognize several persons based on the neural oscillatory hypothesis, by using a recurrent neural network where nonlinear oscillators are used in order to obtain limit-cycle dynamics shown in Fig. 1(b). After pre-processing of feature extraction, we can deal with a complicate recognition problem as a simple combinatorial problem. Our system is, therefore, superior to the existing models in terms of simplicity and flexibility[3][4].

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 395-400

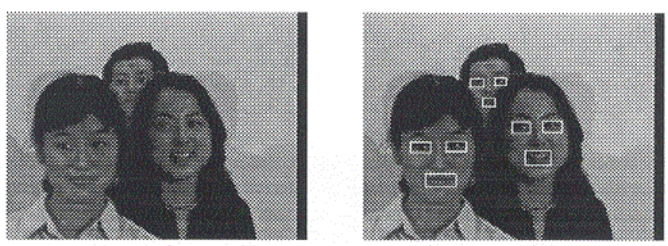


(a) (b)

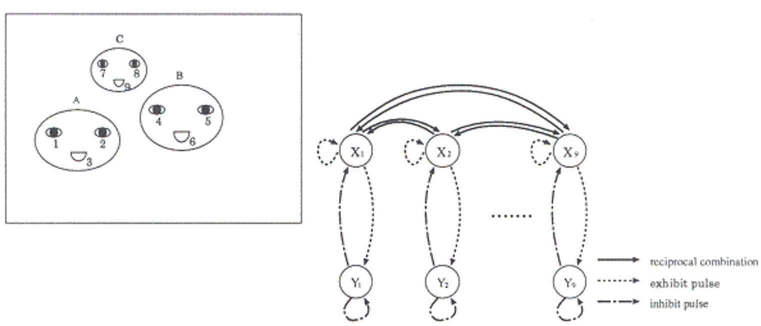
Fig. 1: Objects recognition by figure-ground separation. (a): At times this picture can be identified as a person's face, and at others it can be identified as a pot. (b): Three persons are recognized by figure-ground separation.

2. Our system for grouping face parts

2.1. System Architecture



(a) (b)



(c) (d)

Fig. 2: (a): Target image. (b): Face parts extraction by pre-processing. (c): Input data. (d): Recurrent neural network model.

As shown in Fig. 2(a), (b) and (c), face parts are extracted by pre-processing from a target image. Each part is assigned to a neuron of a recurrent neural network respectively. As mention below, neuron i consists of an excitatory element X_i and an inhibitory element Y_i.

2.2. Grouping of face parts

Grouping is realized by combining applopriate face parts. Since our system has some templates of faces, the result of grouping is compared with the templates.

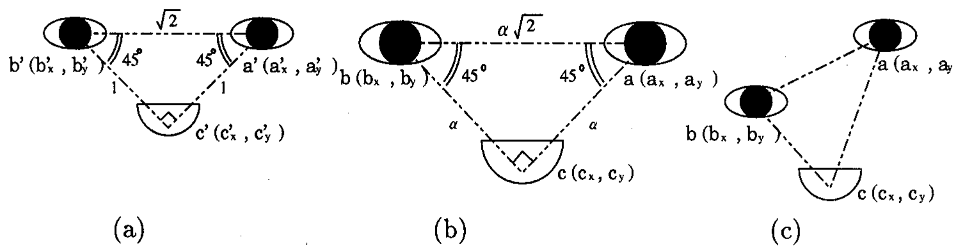


Fig. 3: (a): Face template. (b): Target image which is similar to face template. (c): Target image which is not similar to face template.

For example, suppose that a template is given like Fig. 3 (a), and two results of grouping are shown in Fig. (b) and (c). Let's now define as follows:

$$||\vec{ab}|| = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2} \tag{1}$$

$$cos\angle bac = \frac{||\vec{ab}||^2 + ||\vec{ca}||^2 - ||\vec{bc}||}{2||\vec{ab}||\cdot||\vec{ca}||} \tag{2}$$

For Fig. 3(b), which is similar to Fig. 3(a),

$$|cos\angle b'a'c' - cos\angle bac| = 0, \tag{3}$$

$$\left|\frac{||\vec{a'b'}||}{||\vec{ab}||} - \frac{||\vec{a'c'}||}{||\vec{ac}||}\right| = 0. \tag{4}$$

For Fig. 3(c), on the other hand,

$$|cos\angle b'a'c' - cos\angle bac| > 0, \tag{5}$$

$$\left|\frac{||\vec{a'b'}||}{||\vec{ab}||} - \frac{||\vec{a'c'}||}{||\vec{ac}||}\right| > 0. \tag{6}$$

Grouping is also efficiently promoted by considering the existance probability of parts. Suppose that a right eye and a mouth are grouped in a target image. It is suggested by the corresponding template that a left eye is laid around the point $s(x', y')$ as shown in Fig. 4(a). The distance between $s(x', y')$ and $a(a_x, b_y)$ is described as follows:

$$||\vec{sa}|| = ||\vec{ca}||^2 + \frac{||\vec{bc}||^2}{||\vec{b'c'}||^2}||\vec{c'a'}|| - 2\frac{||\vec{bc}||}{||\vec{b'c'}||}||\vec{ca}||\cdot||\vec{c'a'}||cos(\angle b'c'a' - \angle bca) \tag{7}$$

Above values are independent of an absolute position of each part.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 395-400

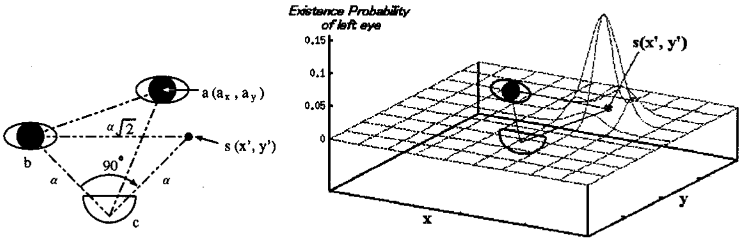
For Fig. 3(b), which is similar to Fig. 3(a),

$$\frac{1}{2\pi}e^{-\frac{1}{2}(\frac{||\vec{sa}||^2}{\sigma^2})} \gg 0, \tag{8}$$

For Fig. 3(c), on the other hand,

$$\frac{1}{2\pi}e^{-\frac{1}{2}(\frac{||\vec{sa}||^2}{\sigma^2})} \approx 0. \tag{9}$$

Existance probability of a part is given based on the Gaussian distribution shown in Fig. 4(b). σ is a constant value which decides the shape of Gaussian distribution.



(a) (b)

Fig. 4: Existance probability of left eye. (a): Estimation of a left eye's position. (b): Existance probability distribution of the ith part as a left eye.

2.3. Neural Netwok Dynamics

Based on a model of DeLiang Wang *et. al*[1], the motion equations of an ith neuron are described as follows:

$$\begin{aligned}\frac{dX_i}{dt} = \varphi X_i^2[& - A\left(\sum_k V_k eye(k) - 2\right) - B\left(\sum_k V_k mouth(k) - 1\right) \\ & - C\sum_a \sum_{b\neq a} \sum_{c\neq a,b} V_a V_b V_c |cos\angle a'b'c' - cos\angle abc| \\ & - D\sum_k \sum_{l\neq k} \sum_m \sum_{n\neq m} V_k V_l V_m V_n \left| \frac{||\vec{kl}||}{||\vec{k'l'}||} - \frac{||\vec{mn}||}{||\vec{m'n'}||} \right| \\ & + E\sum_{p\neq i} \sum_{q\neq i,p} V_p V_q \frac{1}{2\pi} e^{-\frac{1}{2}(\frac{||\vec{si}||^2}{\sigma^2})} \\ & - Ff(Y_i) + G\int_0^t X_i e^{(\tau - t)} d\tau]\end{aligned} \tag{10}$$

$$\frac{dY_i}{dt} = -\frac{Y_i}{\xi} + Hg(-T_{yy}Y_i + T_{yx}X_i) \tag{11}$$

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 395-400

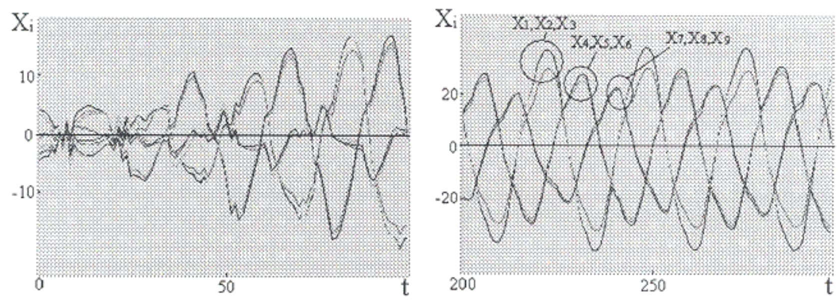
$$f(z) = (1-\mu)z + \mu z^2, \quad g(z) = \frac{1}{1+e^{\frac{-(z-\theta)}{\lambda}}}, \quad V_i = \begin{cases} 1 & \text{if } X_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$eye(i) = \begin{cases} 1 & \text{if } i\text{th part is an eye} \\ 0 & \text{otherwise} \end{cases}, \; mouth(i) = \begin{cases} 1 & \text{if } i\text{th part is a mouth} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

$\varphi, A, B, C, D, E, F, G, H, \xi, T_{yy}, T_{yx}, \mu, \theta$ and λ are constant values.

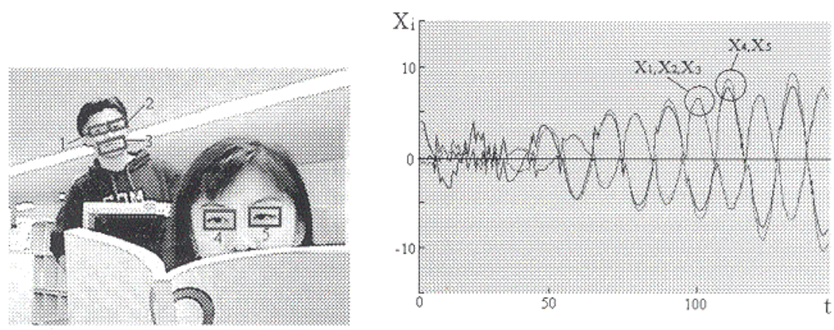
3. Simulation Results

A simulation result using Fig. 2(c) is shown in Fig. 5. A limit-cycle dynamics of nine neurons is observed from about 200 iteration steps. Since three group of neurons are found, three persons are descriminated intermittently.



(a) (b)

Fig. 5: Simulation result. Target data is Fig. 2. The person A's right eye, left eye, and mouth are assigned to X_1, X_2, and X_3. The person B's right eye, left eye, and mouth are assigned to X_4, X_5, and X_6. The person C's right eye, left eye, and mouth are assigned to X_7, X_8, and X_9. (a): $0 \leq t \leq 100$, and (b): $200 \leq t \leq 300$.



(a) (b)

Fig. 6: (a): Target image where a woman's mouth is lost. (b): Dynamics of five neurons ($0 \leq t \leq 150$).

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 395-400

Fig. 6 shows another simulation result. Even though a mouth of a woman in the front of the picture is lost, the system can detect two persons.

4. Discussion

The above simulation results show that our system can detect some persons one after another. One of advantages of the system is to realize simply an image normalization problem on size, rotation and shift of an face. Even though a face part is lost, the behavior of neurons are converged to the limit-cycle(See Fig. 6). If the pre-processing extracts such feature parts, nose, ears, contours and hair, we can add those constraint conditions to the motion equation(Eq.(10)).

5. Conclusion

As the result of transforming a complex human detection problem to a simple combinatorial problem, our system has only to consider how to group several face parts even though some of the parts are lost. The system is superior to others in terms of simplicity and flexibility[3][4].

References

1. Raphael Ritz , Wulfram Gerstner, and J. Leo van Hemmen, "Associative Binding and Segregation in a Network of Spiking Neurons", Models of Neural Networks II, Chapter 5, Springer-Verlag, 1996.
2. Patricia S. Churchland and Terrence J. Sejnowski, "THE COMPUTATIONAL BRAIN", Chapter 2, MIT Press, 1996.
3. DeLiang Wang, Joachim Buhmann, "Pattern Segmentation in Associative Momory", Neural Computation 2, pp.94-106, 1990.
4. Yutaka Hirakura, Yoko Yamaguchi, Hiroshi Shimizu, and Shinichi Nagai,"Dynamic Linking Among Neural Oscillators Leads to Flexible Pattern Recognition with Figure-Ground Separation", Neural Networks, vol. 9, no. 2, pp.189-209, 1996.
5. Yoko Yamaguchi and Hiroshi Shimizu, "Pattern Recognition With Figure-Ground Separation by Generation of Coherent Oscillations", Neural Networks, vol. 7, no. 1, pp.49-63, 1994. Neural Networks, Amsterdam, Sep. 13-16, 1993.
6. David Horn, Irit Opher, "Temporal Segmentation in a Neural Dynamic System", Neural Computation 8, pp.373-389, 1996.