PERGAMON

Engineering Applications of Artificial Intelligence 13 (2000) 543–548

Engineering Applications of
ARTIFICIAL
INTELLIGENCE

www.elsevier.com/locate/engappai

The third eye approach to innovative designs and applications into the 21st century: human recognition system by nonlinear oscillations

S. Oka*, Y. Takefuji, W. Huang

Graduate School of Media and Governance, Keio University, Kanagawa, Japan

Abstract

The field of remote sensing and sensor technology has undergone tremendous development in the past decades. Sensors technologies of all kinds such as electro-optics, acoustic, active/passive UV to LWIR, ground-penetrating radar, passive mm wavelength, X-ray tomography, neutron activation imaging, multi-spectral, hyper-spectral, and ultra-spectral imaging, will provide valuable images that normal CCD camera cannot offer. By combining algorithms and images taken by sensors at different part of the electromagnetic spectrum, we will be able to extract valuable images automatically. By using multi-spectral images and processing them with neural network computing, our "Third Eye" team is able to extract human face features from those images. In this paper, we will present an application for detecting human facial parts, images taken by different imaging systems and sensors, and the current status of image processing applications. © 2000 Published by Elsevier Science Ltd.

Keywords: Hyper-spectral image; Neural network; Human recognition; Nonlinear oscillation

1. Introduction

Light is a form of electromagnetic radiation. Other forms of electromagnetic radiation include radio waves, microwaves, infrared radiation, ultraviolet rays, X-rays, and γ-rays. All of these are known collectively as the electromagnetic spectrum. The rainbow of colors that we see in visible light represents only a very small portion of the electromagnetic spectrum. The electromagnetic spectrum ranges from γ-rays to radio waves with everything else in between. The reflectance spectra of most materials on the Earth's surface contain characteristics or diagnostic absorption features. Remote sensors, such as hyper-spectral imager developed by TRW, capable of acquiring complete reflectance spectra over large areas offer a powerful tool for study of the Earth and the environment (http://www.trw.com/). With a CCD camera, we are only able to see the visible part of the electromagnetic spectrum that ranges from 0.4 to 0.7 μm of the spectrum, thus, any range below or beyond the visible part of the EMS are invisible to human eyes (see Fig. 1). However, by combining images taken at different part of the spectrum and processing them, we will be able to reveal many valuable images. Curtiss Davis at NRL defines that multi-spectral images are captured by several wide bands, hyper-spectral images more than 60 narrow bands with less than 10 nm, and ultra-spectral images more than 100 very narrow bands with less than 0.1 nm (http://www.nrl.navy.mil/). Unique algorithms used in conjunction with different imaging systems and sensors will allow exploration of a large variety of basic problems in fields such as Earth science, vegetation studies, geology, semiconductor manufacturing, biology, biological imaging, material processing, environment, medical imaging, ground target detection, chemical identification, and inspection (http://www.iinspect.com/; http://www.certified-testing.com/).

* Corresponding author.

0952-1976/00/$ - see front matter © 2000 Published by Elsevier Science Ltd.
PII: S0952-1976(00)00034-8

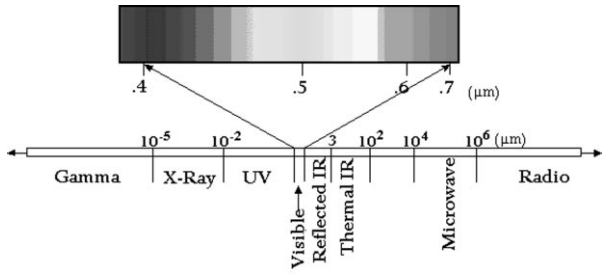

Fig. 1. Hyper-spectrum wave length.

In the field of machine vision and visual inspection, a company called SRI International has pioneered, designed, and developed an automatic visual inspection system that has applied widely in inspection of defects, cracks, or structural failures in different area such as on NASA shuttle tile inspection (http://www.erg.sri.com/). There are over 20,000 thermal tiles that cover the outer surfaces of the shuttle. Each of these tiles is life-critical because a failure can result in the shuttle's aluminum skin overheating on reentry, thus, causing a catastrophic system failure. Consequently, each tile is manually inspected after each flight to ensure that there are no cracks, dents, or other structural failures that require rework. To reduce human error and inspection time, SRI International has developed an imaging system and inspection algorithm to detect damages to the thermal tiles. This system demonstrates the fact that by building or manipulating different sensors and developing specific algorithms, we will be able to produce incredible images which will facilitate us in defect inspection or manufacturing that will ultimately save time, money, and human lives. Recently, our team developed a new visualization algorithm for special image capturing sensors in various fields (http://www.neuro.sfc.keio.ac.jp/). As shown in Figs. 2 and 4, the visualization algorithm shows the land utilization from remote sensing images, and extracts the malignant tumor from MRI. Our project is supported by NHK and NAL (see Figs. 3 and 5) (http://www.nal.go.jp/).

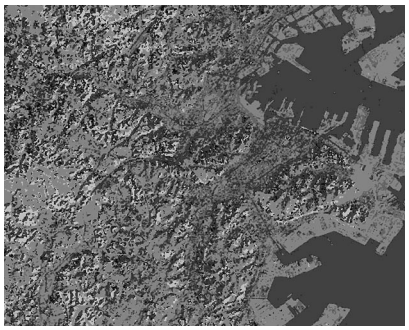

Fig. 2. Remote sensing image by LANDSAT-TM.

2. Human counting system

We show a neural network application of the third eye project — Human Counting System by using a CCD camera and an infrared camera. This system detects several faces in a scene simultaneously through two phases: extracting and grouping facial parts. After facial parts including eyes and mouths are extracted from a target image in the extracting phase, they are fed to a recurrent neural network composed of non-linear oscillatory components (Ohkita et al., 1998). Any pair of neurons in this network is synchronized, respectively, if each of them represents a facial part of the same face. The proposed neural network can detect several faces one after another by composing facial parts extracted from the extracting phase. One of the advantages in this neural network is for a real-time system for counting the number of people.

3. Neural network dynamics

Grouping is realized by combining the appropriate facial parts. A neuron is assigned to each extracted facial part. Since our system has some templates of faces, the result of the grouping is compared with them. Suppose that a template is given in Fig. 6(a), and a result of grouping is shown in Fig. 6(b). Based on the similarity between a template and a result of grouping, the motion equation of an ith neuron is shown in Eq. (1), where N is the number of neurons. The neuron model we use is the McCulloch–Pitts binary neuron model. ξ, ψ, ζ, A, B, C, D, E, F, G, t', σ, μ, λ, ϕ and θ are constant values.

$$\begin{aligned}\frac{\mathrm{d}X_i}{\mathrm{d}t} = \xi\left(\psi - \varsigma X_i^2\right)\Bigg[& - A\left(\sum_k^N V_k \mathrm{eye}(k) - 2\right) \\ & - B\left(\sum_k^N V_k \mathrm{mouth}(k) - 1\right) \\ & - C\sum_a^N \sum_{b\neq a}^N \sum_{c\neq a,\, b}^N V_a V_b V_c \left|\cos\angle a'b'c' - \cos abc\right| \\ & - D\sum_k^N \sum_{l\neq k}^N \sum_m^N \sum_{n\neq m}^N V_k V_l V_m V_n \left|\frac{\|\vec{kl}\|}{\|\vec{k'l'}\|} - \frac{\|\vec{mn}\|}{\|\vec{m'n'}\|}\right| \\ & + E\sum_{p\neq i}^N \sum_{q\neq i,\, p}^N V_p V_q \frac{1}{2\pi}\exp\left(-\|\vec{si}\|^2/2\sigma^2\right) \\ & - F\sum_{j\neq i}^N V_j\left(u(i,j) - \phi\right) \\ & - G\int_{t-t'}^t X_i(\tau)\frac{1}{1+\exp\left(-(\tau-\mu)/\lambda\right)}\mathrm{d}\tau\Bigg] \end{aligned} \tag{1}$$

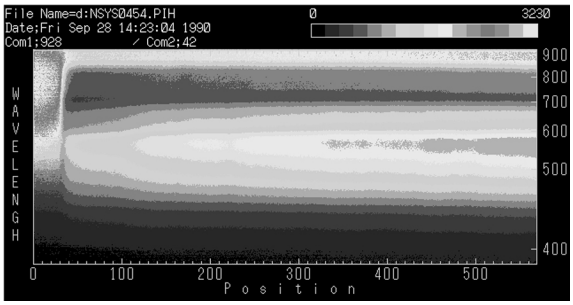

Fig. 3. Hyper-spectrum sensing for detecting sludge in ocean.

Grouping is efficiently promoted by considering the existence probability of parts. Suppose that the right eye and the mouth are grouped in a target image. The corresponding template suggests that a left eye is laid around the point $s(x', y')$ as shown in Fig. 6(c). The distance between $s(x', y')$ and $a(a_x, b_y)$ is described as follows:

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

The existence probability of a part given is based on the Gaussian distribution. σ is a constant value which determines the shape of the Gaussian distribution. In addition, we use an infrared image in order to realize more precise grouping. Infrared images can detect the temperature of the human skin as shown in Fig. 7. The temperature information is useful for precise grouping. In the Eq. (1), u(a, b) is defined as follows:

$$u(a, b) = \begin{cases} \int_{a_x}^{b_x} r\left(\mathrm{IR}\left(x\frac{b_y - a_y}{b_x - a_x}\right)(x - a_x) + a_y\right)\mathrm{d}x & \text{if } a_x \neq b_x \\ \int_{a_x}^{b_x} r\big(\mathrm{IR}(a_x, y)\big)\mathrm{d}y & \text{otherwise} \end{cases} \quad (3)$$

3.1. Simulation-1

A sample target image and a simulation result are shown in Fig. 7. Person A's right eye, left eye, and mouth are assigned to X_1, X_2, and X_3. Person B's right eye, left eye, and mouth are assigned to X_4, X_5, and X_6. Person C's right eye, left eye, and mouth are assigned to X_7, X_8, and X_9. Nine neurons converge to the limit-cycle, respectively, where the behaviors of nine neurons are very coherent. A limit-cycle dynamics of nine neurons can be observed from about 50 iteration steps. Since three groups of neurons are found, three people are discriminated intermittently. The phase difference of three oscillation groups are about $2\pi/3$, respectively. The period of the oscillations is about 30 iteration steps, which is almost the same for all nine neurons.

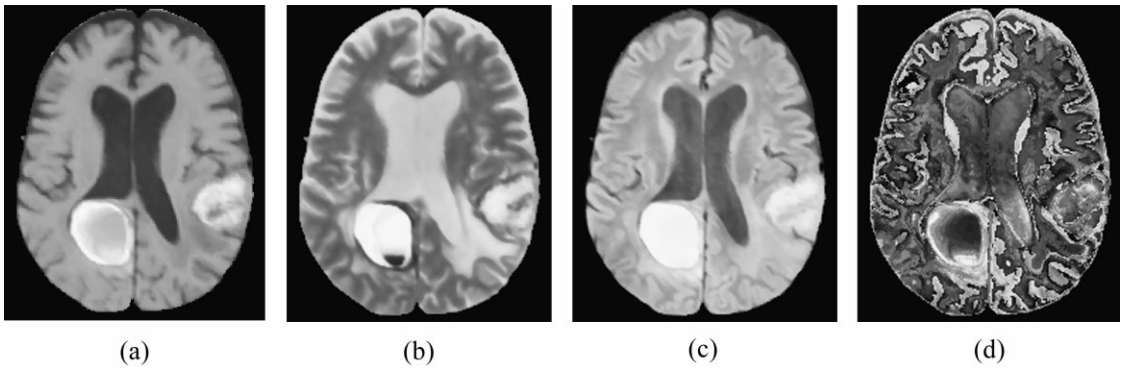

Fig. 4. MRI analysis: (a) T1 image, (b) T2 image, (c) proton image, and (d) coloring result.

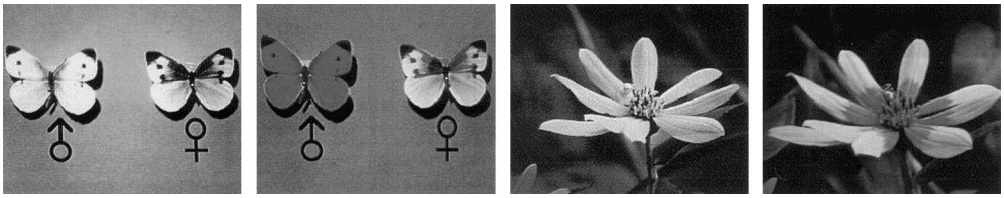

Fig. 5. CCD camera image and ultra-violet capturing by BeeCam.

$$\left\|\overrightarrow{sa}\right\| = \left\|\overrightarrow{ca}\right\|^2 + \frac{\left\|\overrightarrow{bc}\right\|^2}{\left\|\overrightarrow{b'c'}\right\|^2}\left\|\overrightarrow{c'a'}\right\|^2 - 2\frac{\left\|\overrightarrow{bc}\right\|}{\left\|\overrightarrow{b'c'}\right\|}\left\|\overrightarrow{ca}\right\| \cdot \left\|\overrightarrow{c'a'}\right\| \cos(\angle b'c'a' - \angle bca) \quad 2.$$

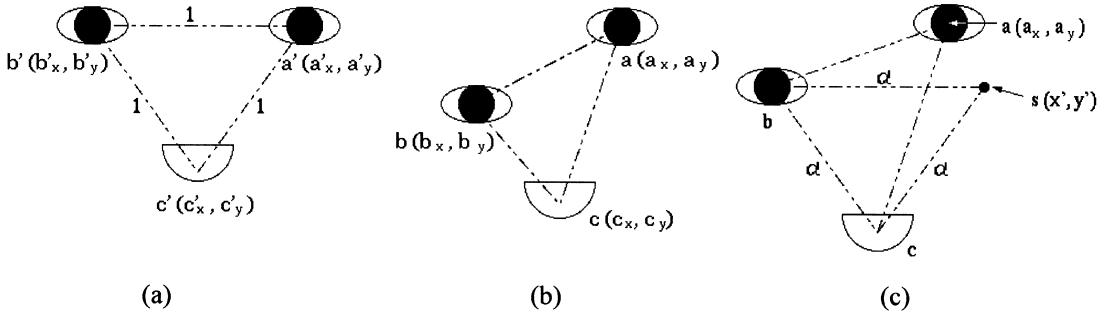

Fig. 6. (a) Facial template, (b) a result of grouping, (c) positional gap between the template and the grouping.

This system is implemented in C language on SUN Ultra2 (2 × UltraSPARC 168 MHz). The computation time to converge to the limitcycle is about 1.0 s. The size of the tested data is 296 × 222 pixel, and 8-bit gray scaled. The constant values are set as follows for all the simulations in this paper.

$\xi = 0.2$, $\psi = 0.8$, $\zeta = 2.0 \times 10^{-8}$, $A = 3.5$, $B = 1.0$,

$C = 0.1$, $D = 0.01$, $E = 8.0$, $F = 10.0$, $G = 0.25$,

$t' = 50$, $\sigma = 1.3$, $\mu = 25$, $\lambda = 10$, $\phi = 0.5$, and $\theta = 30$.

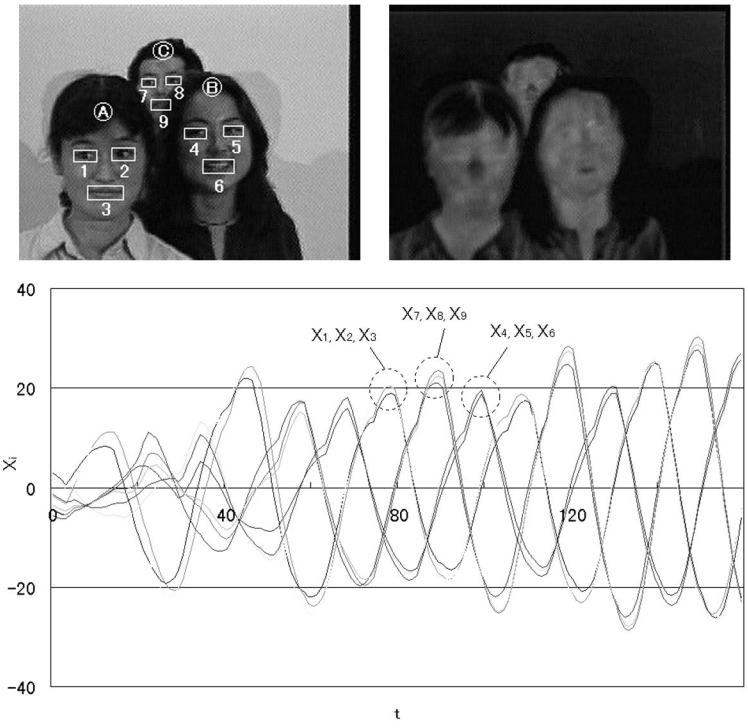

Fig. 7. Target image and the neural network dynamics of simulation-1.

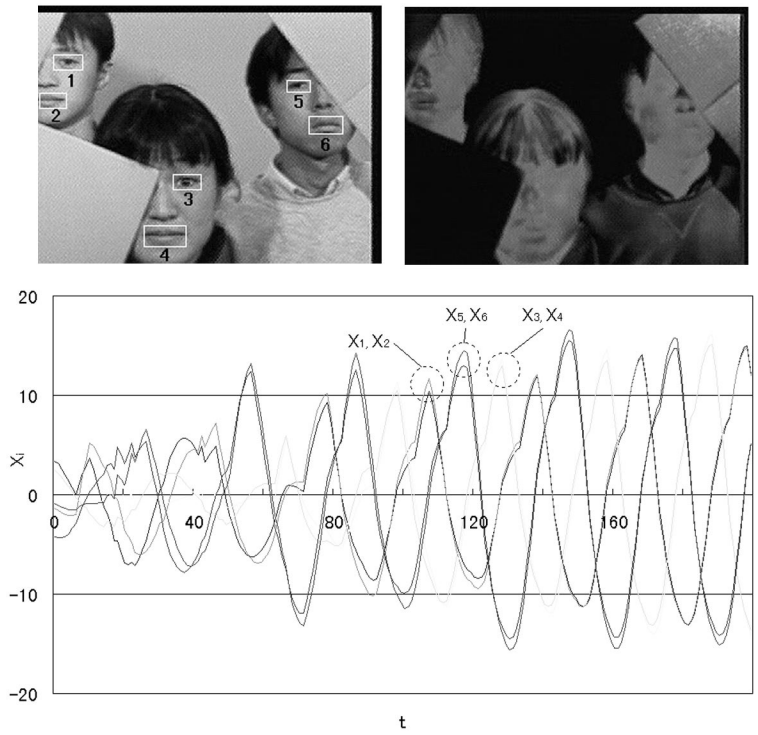

Fig. 8. Target image and the neural network dynamics of simulation-2.

3.2. Simulation-2

Another target image and a simulation result are shown in Fig. 8. A limit-cycle dynamics of six neurons is observed from about 100 iteration steps. The system succeeds in grouping three people, even though the eyes of the man on the left, the woman in the center, and the man on the right are lost, respectively.

4. Discussion

Feature grouping is realized by the simultaneous activation of neurons which encode a same face (Wang et al., 1990; Hirakura et al., 1996). A face consists of a set of several parts, and they are chosen by comparing them with a facial template. If some parts are lost, other parts have to make up for the lost information.

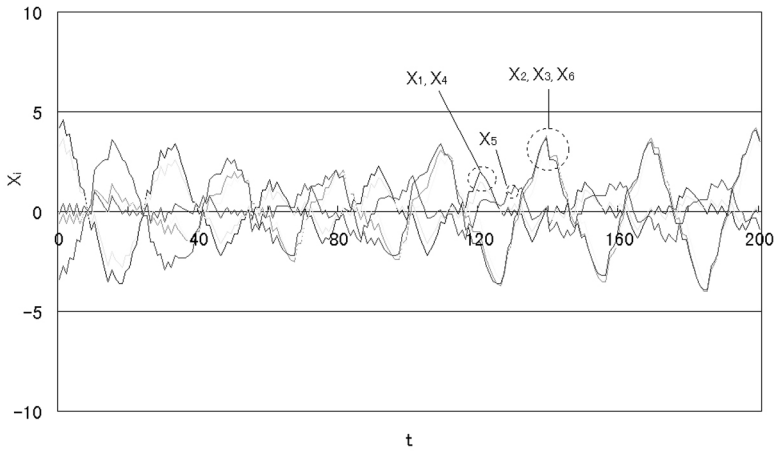

Fig. 9. The neural network dynamics of simulation-2 without the infrared image.

The system needs at least three parts to detect a person. A simulation result using the target image in Fig. 8 without the infrared image is shown in Fig. 9, where the system cannot detect three persons correctly because the information is not sufficient for grouping. If some parts are lost in a target image, the system will not work well without an infrared image. In other words, an infrared image helps the system to obtain precise grouping when some parts are lost. If the first phase extracts such feature parts, nose, ears, contours and hair, we can add those constraint conditions to the motion equation (Eq. (1)) in order to obtain a more precise grouping.

5. Conclusion

As a result of transforming a complex human detection problem to a simple combinatorial problem, our system only has to consider how to group several face parts even when some of the parts are lost. The number of persons the system can identify simultaneously is up to four (Horn and Opher, 1996; Ritz et al., 1996).

References

Horn, D., Opher, I., 1996. Temporal segmentation in a neural dynamic system. Neural Computation 8, 373–389.

Wang, D., Buhmann, J., von der Malsburg, C., 1990. Pattern segmentation in associative memory. Neural Computation 2, 94–106.

Ritz, R., Gerstner, W., van Hemmen, J.L., 1996. Associative Binding and Segregation in a Network of Spiking Neurons Models of Neural Networks II. Springer–Verlag, Berlin.

Ohkita, S., Ajioka, Y., Takefuji, Y., 1998. Detection of human facial parts using infrared and visible Image. In: Proc. of EANN98 in Gibraltar, 54–57.

Hirakura, Y., Yamaguchi, Y., Nagai, S., 1996. Dynamic linking among neural oscillators leads to flexible pattern recognition with figure-ground separation. Neural Networks 9 2, 189–209.

http://www.neuro.sfc.keio.ac.jp/.

http://www.nrl.navy.mil/.

http://www.nal.go.jp/.

http://www.iinspect.com/.

http://www.certified-testing.com/.

http://www.trw.com/.

http://www.erg.sri.com/.

Souichi Oka is a doctorate fellow of Keio University. He is at present working as chief scientist in Neural Research Laboratory. His research interests include Neural computing and hyper-spectral computing.

Yoshiyasu Takefuji is working as Professor of Keio University since 1992. He is also Government advisor to NCC of Philippines, VITTI of Vietnam, Jordan IT, Multi-media University in Malaysia. His research interests include Neural computing and hyper-spectral computing.

William Huang has done his masters from Keio University. His research interests include Neural computing and electrical engineering.