A Face/Object Recognition System Using FPGA Implementation of Coarse Region Segmentation

T. Nakano[1], T. Morie[1], and A. Iwata[2]

1 Kyushu Institute of Technology, Kitakyushu, 808-0196 Japan
2 Hiroshima University, Higashi-Hiroshima, 739-8526 Japan
nakano-teppei@edu.brain.kyutech.ac.jp

Abstract: A PC system for recognition of natural scene images including human faces and various objects is proposed. Coarse region segmentation of real images with 64×64 pixels at the video rate is achieved by using the FPGA implementation of resistive-fuse networks. A flexible template matching based on dynamic-link architecture is performed on our PC system.

Keywords: face recognition, object recognition, resistive-fuse, FPGA, Gabor wavelet transform, dynamic-link matching

1. Introduction

In order to recognize the image of a natural scene including several objects, meaningful image regions should be segmented, extracted and recognized separately for reducing the complexity of the problem. Our proposed recognition procedure consists of coarse region segmentation/extraction, Gabor wavelet transformation (GWT) and dynamic-link matching.

The resistive-fuse network is a well-known image segmentation processing model in which image edges are preserved and noise is eliminated [1]. Some attempts for its analog LSI implementation have been proposed [2, 3]. However, practical design for large-scale analog resistive-fuse network circuits (more than 100×100 pixels) is very difficult because of unexpected parasitic components and various non-idealities in analog circuits. Accordingly, we have proposed the resistive-fuse network circuit using pulse-width-modulation cellular neural networks (PWM-CNN) [4, 5], and have demonstrated successful LSI implementation for 1-D case [6]. We have also applied the resistive-fuse network model to digital image processing [7].

In this paper, we propose a semi-realtime recognition system using a PC with an FPGA board. In Sec. 2, we describe the face/object recognition system. In Sec. 3, we present our system composed of PC/FPGA and show face and object recognition results.

2. Face/Object Recognition System

2.1 Processing Flow for Face/Object Recognition

Figure 1 shows the processing flow for face/object recognition from natural scene images [8]. First, the edge of each object is extracted by using coarse region segmentation. The coarse region segmentation performed by a resistive-fuse network is a unique process in our system. Then, the posi-

tion and the shape of each region are determined one by one using the conventional labeling process. Using this information, the image of the region is extracted from the input image and GWT is performed. Next, template matching using GWT based on dynamic-link architecture [9] is performed. This architecture is known to be robust to image distortion including a change in face expression, direction and rotation, because it searches the best matched stored image by moving corresponding sampling points for matching between input and stored images.

Section 2.2 describes the coarse region segmentation using the resistive-fuse network circuit. Section 2.3 describes feature extraction using GWT, and Sec. 2.4 describes flexible template matching using dynamic-link matching.

2.2 Coarse Region Segmentation Using Resistive-Fuse Networks

Figure 2(a) shows an original analog circuit of the resistive-fuse network, in which nonlinear resistance elements connect the neighboring pixel nodes for image region segmentation. The image input I_n for pixel n is given as a voltage source. The processing result (output) O_n is given as
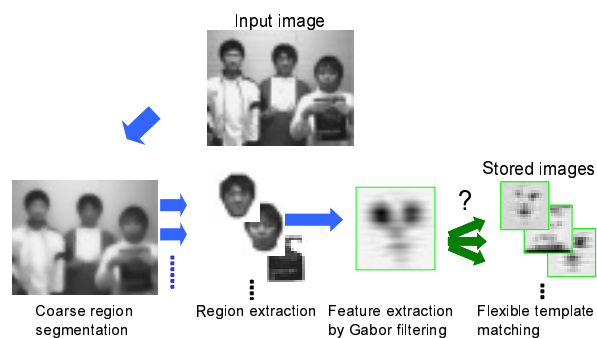

Figure 1: Processing flow for face/object recognition from natural scene images.

a node voltage. If the difference between neighboring pixels $|O_n - O_k|$ is smaller than the threshold value δ, the network behaves as a simple linear resistive network, and image smoothing is performed. On the other hand, if $|O_n - O_k| \geq \delta$, the pixel nodes are disconnected each other, and such pixels are recognized as an image edge.

Moreover, by changing I-V characteristics of the nonlinear resistance element from linear resistance to resistive-fuse, as shown in Fig. 2(b), coarse region segmentation can be achieved. In this processing, for example, a whole face region is segmented irrespective of small facial parts such as brows, eyes, a nose, and a mouth, as shown in Fig. 2(c).

We propose a digital circuit that emulates the operation of the analog nonlinear circuit by discrete-time dynamics based on clock operation. The change in each node voltage is calculated using Kirchhoff's current law, and the steady state is obtained by repeating the updating process.

Figure 3(a) shows the memory assignment in the whole image, and the digital resistive-fuse circuit is shown in Figure 3(b). The look up table memory LUT_1 transforms the input data x to output data y according to linear function: $y = \sigma x$. The look up table memory LUT_2 corresponds to the nonlinear function $G(\cdot)$ shown in Figure 2(b). The pixel data I_n is stored in the "Source memory," and the pixel data O_n is stored in the "Destination memory." The detail of the digital LSI implementation of resistive-fuse networks is described in Ref. [7].

The proposed algorithm was implemented using an FPGA. The FPGA used is ALTERA EP20K400EBC652-1X, which is included in a PCI board shown in Figure 4. When the clock frequency was 40 MHz, the processing time was less than 20 ms for an image of 64×64 pixels.

2.3 Feature Extraction Using GWT

The GWT is performed for a selected region image. The convolution kernel for GWT can be given by

$$\psi_{\vec{\omega_0}}(\vec{x}) = \frac{1}{2\sqrt{\pi}\sigma} \exp\left(-\frac{\vec{\omega_0^2}\vec{x^2}}{\sigma^2} + j\vec{\omega_0}\vec{x}\right), \qquad (1)$$

where $\vec{\omega_0}$ is a parameter that determines the frequency and the direction. Figure 5 shows an example of GWTs for a
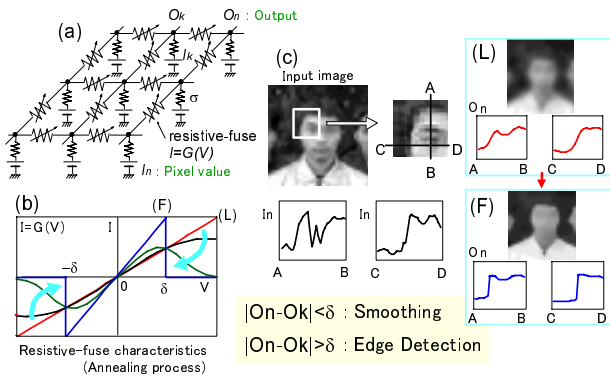
human face. Our recognition system uses GWTs with four directions and five frequencies.

2.4 Dynamic-Link Matching

The matching process has two phases; a memory phase and a recognition phase. In the memory phase, GWT coefficients at all pixels in the extracted image region are stored in the memory, while in the recognition phase GWT coefficients at only sampling points (8×8) of the image are used for the flexible template matching process described below. Therefore, the processing time for GWT in the recognition phase is much shorter than that in the memory phase, and is negligible in the total recognition time.

The matching evaluation is done by a trade-off between better matching in Gabor features and less distortion in the sampling points (Figure 6). The evaluation function is defined as:

$$E = E_v - \lambda E_e. \qquad (2)$$

Here, E_v expresses the difference between GWT coefficients of the input image and those of stored images;

$$E_v = \sum_{v \in V} \frac{\overrightarrow{J_v^I} \cdot \overrightarrow{J_v^S}}{\|J_v^I\| \, \|J_v^S\|}, \qquad (3)$$

where V is a set of sampling points; J_v is a GWT coefficient vector at point v; superscripts I and S indicate the input im-
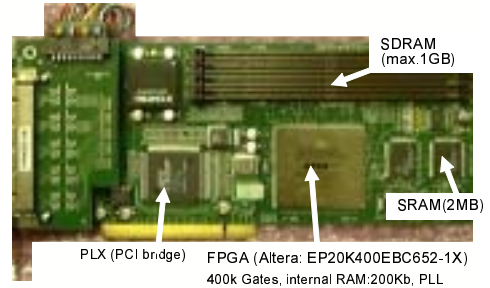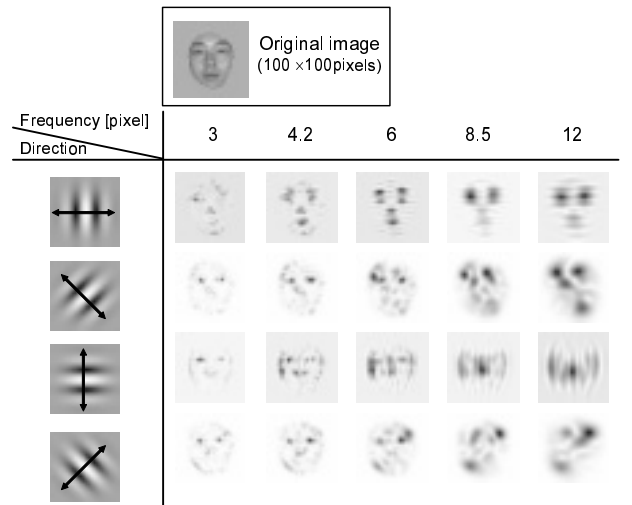

Figure 4: PCI FPGA board.

Figure 5: Example of GWTs for a human face.

Figure 2: Resistive-fuse network circuit and the principle of coarse image region segmentation using it.
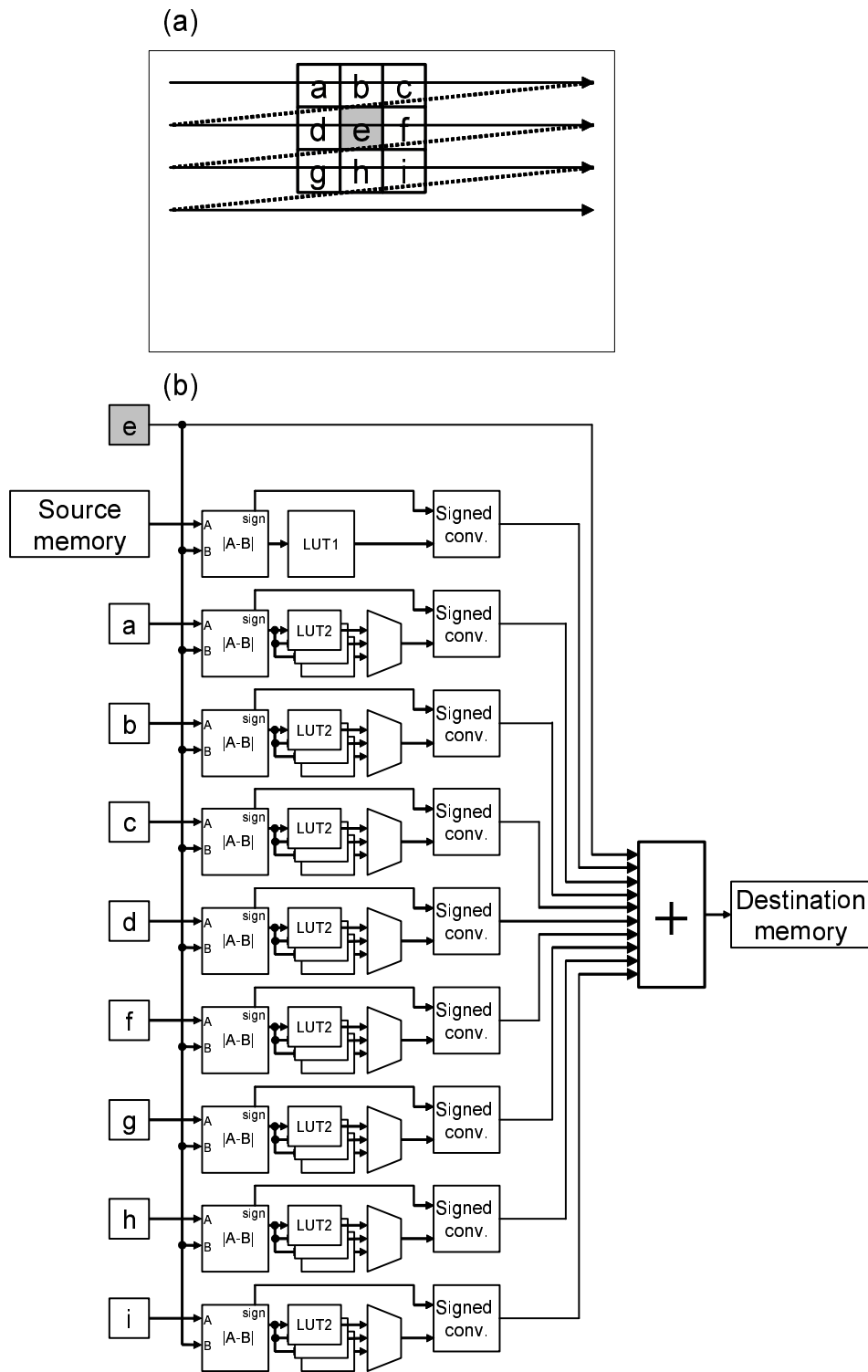
Figure 3: Digital LSI implementation of resistive-fuse networks: (a) assigned 3×3 pixels in the whole image pixels and (b) resistive-fuse circuit.

age and the stored image, respectively. On the other hand, E_e expresses the distortion of the sampling point lattice;

$$E_e = \sum_{v \in V} \sum_{w \in N_v} \left| D_{vw}^I - D_{vw}^S \right| \qquad (4)$$

where N_v is a set of neighboring matching points of v; D_{vw} is the distance between v and w.

Our system uses a *steepest descent method* for searching the matching points having the minimum E. Here, outermost sampling points are fixed. The flowchart of the matching algorithm is shown in Fig. 7.

3. Face/Object Recognition System Composed of PC/FPGA and Its Demonstration

Figure 8 shows our recognition system, which consists of a camera and a PC including an image capture board and a PCI board with an FPGA. Table 1 shows the specification of the PC and the software development environment. The p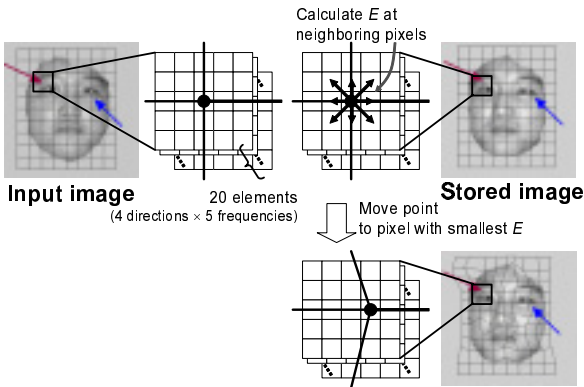rocessing flow for recognition is described below. An image of 320×240 pixels is captured by the image capture board. Coarse region segmentation of the image with a size reduced to 64×64 pixels is performed by the PCI FPGA board. Then, using the edges of the segmented regions, the region images are extracted one by one using the conventional labeling process. The size of the images is adjusted to 100×100 pixels. Finally, the flexible template matching is performed.
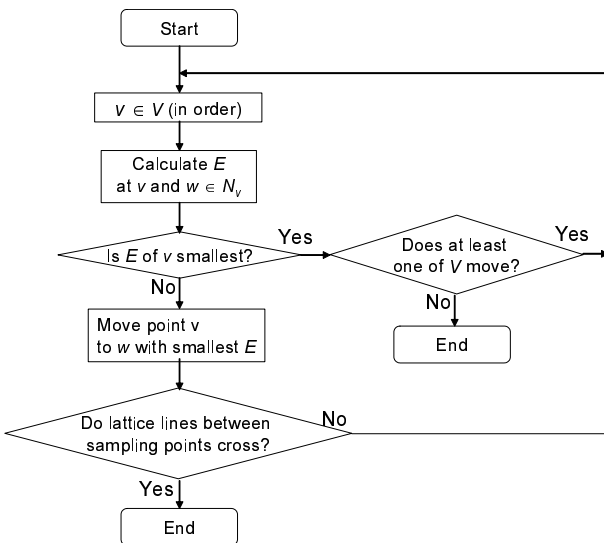
Figure 9 shows display images of the PC. Figure 9(I) demonstrates a human face recognition result and (II) demonstrates a book recognition result. An input image is shown in Figure 9(a), which is a snapshot of a person standing in front of a window-shade. In the usual edge detection, detail edges of the face and the shade are detected as shown in Fig. 9(b); No meaningful region can be segmented. In contrast, the resistive-fuse processing result (c) and its edge detection result (d) show that only the whole human face region is successfully segmented. Segmented regions extracted are listed at (e). The regions of extracted images are indicated by rectangles in (a). One of them is chosen (f) and compared with the stored images listed at (g). A numeral under each stored image is the value of evaluation function E. The best matching image is shown in (h), which indicates a correct recognition result.

Figure 10 shows examples of matching point lattices. Point (i) and (i') identify the right eye correctly by searching smallest E using the dynamic-link matching.

4. Conclusions

The face/object recognition system using coarse region segmentation and flexible template matching was presented. The resistive-fuse network circuit was implemented in an FPGA by a pixel serial approach, and coarse region segmentation of real images with 64×64 pixels at the video rate was

Figure 6: Principle of flexible template matching.

Figure 7: Flowchart of flexible template matching.
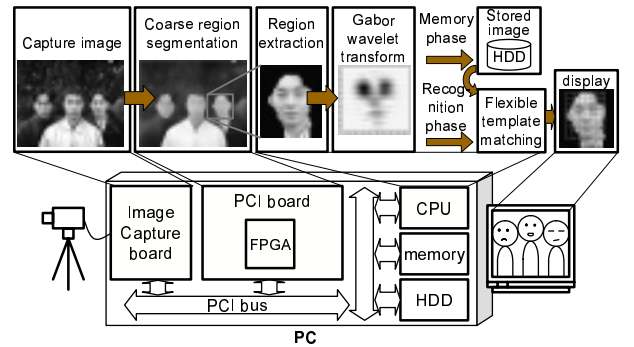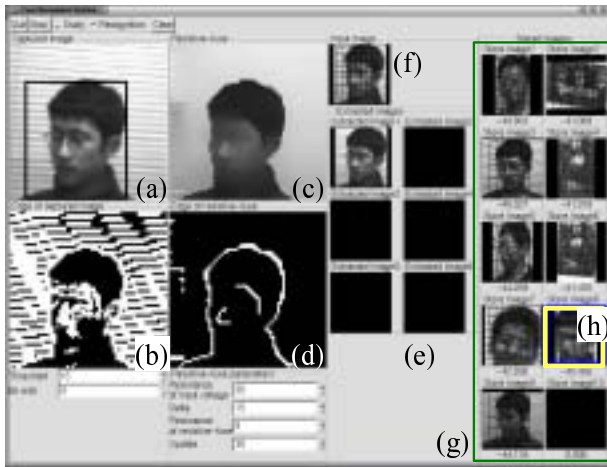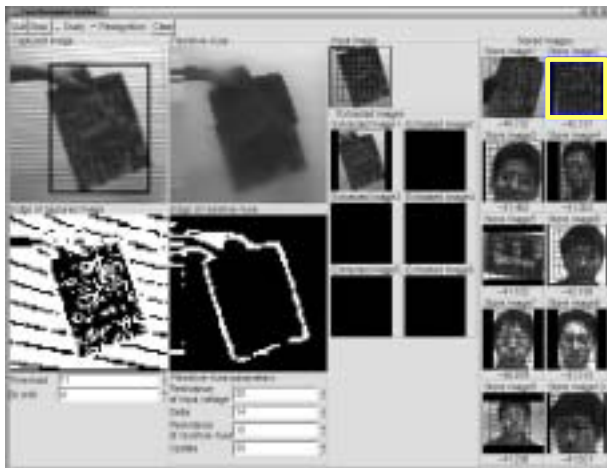

Figure 8: Face/object recognition system.

Table 1: PC specification and software development environment.

CPU	Intel Pentium 4 / 1.8GHz
Main memory	1 Gbytes
OS	Linux (2.2.18)
GUI toolkit	GTK (1.2.10)
Programming language	C (egcs 2.91.66)

achieved. The flexible template matching using dynamic-link architecture was performed in the PC system.

(I) Human face recognition.

(II) Book recognition.

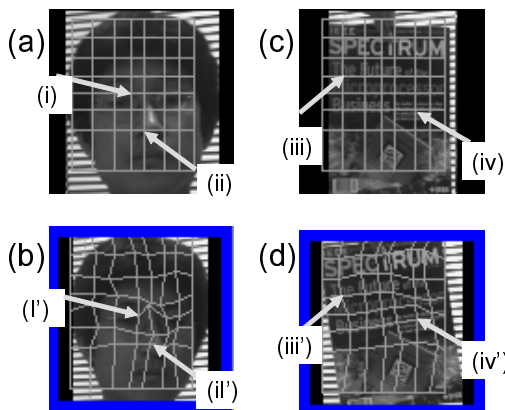Figure 9: Demonstration of our recognition system.

Figure 10: Examples of matching point lattices: (a) and (c) are input images, and (b) and (d) are stored images; (i)-(i'), (ii)-(ii'), (iii)-(iii'), and (iv)-(iv') are the corresponding points, respectively.

References

[1] J. Harris, C. Koch, and J. Luo, "Resistive fuses: Analog hardware for detecting discontinuities in early vision," in Analog VLSI Implementation of Neural Systems, ed. C. Mead and M. Ismail, pp.27–55, Kluwer Academic Publishers, 1989.

[2] P.C. Yu, S.J. Decker, H.S. Lee, C.G. Sodini, and J.L. Wyatt, Jr., "CMOS resistive fuses for image smoothing and segmentation," IEEE J. Solid-State Circuits, vol.27, pp.545–553, 1992.

[3] T. Sawaji, T. Sakai, H. Nagai, and T. Matsumoto, "A floating-gate MOS implementation of resistive fuse," Neural Computation, vol.10, no.2, pp.485–498, 1998.

[4] T. Morie, M. Miyake, S. Nishijima, M. Nagata, and A. Iwata, "A multi-functional cellular neural network circuit using pulse modulation signals for image recognition," Proc. Int. Conf. on Neural Information Processing (ICONIP), Taejon, Korea, pp.613–617, Nov. 2000.

[5] H. Ando, T. Morie, M. Miyake, M. Nagata, and A. Iwata, "Image segmentation/extraction using nonlinear cellular networks and their VLSI implementation using pulse-modulation techniques," IEICE Trans. Fundamentals., vol.E85-A, no.2, pp.381–388, 2002.

[6] T. Morie, M. Miyake, M. Nagata, and A. Iwata, "A 1-D CMOS PWM cellular neural network circuit and resistive-fuse network operation," Ext. Abs. of Int. Conf. on Solid State Devices and Materials (SSDM), Tokyo, pp.90–91, Sept. 2001.

[7] T. Nakano, H. Ando, H. Ishizu, T. Morie, and A. Iwata, "Coarse image region segmentation using resistive-fuse networks implemented in fpga." to appear in 7th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2003), Orlando, July 27-30, 2003.

[8] T. Morie, M. Nagata, and A. Iwata, "Design of a pixel-parallel feature extraction VLSI system for biologically-inspired object recognition methods," Proc. Int. Symp. on Nonlinear Theory and its Applications (NOLTA2001), Zao, Japan, pp.371–374, Oct. 2001.

[9] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R.P. Würtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," IEEE Trans. Comput., vol.42, no.3, pp.300–311, 1993.