Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Image Processing for a Tactile/Vision Substitution System Using Digital CNN

Chien-Nan Lin, Sung-Nien Yu, and Jin-Cheng Hu
cnlin@samlab.ee.ccu.edu.tw, yusn@ee.ccu.edu.tw, macoto_hu@baycom.com.tw
Department of Electrical Engineering, National Chung Cheng University, Taiwan

***Abstract*—In view of the parallel processing and easy implementation properties of CNN, we propose to use digital CNN as the image processor of a tactile/vision substitution system (TVSS). The digital CNN processor is used to execute the wavelet down-sampling filtering and the half-toning operations, aiming to extract important features from the images. A template combination method is used to embed the two image processing functions into a single CNN processor. The digital CNN processor is implemented on an intellectual property (IP) and is implemented on a XILINX VIRTEX II 2000 FPGA board. Experiments are designated to test the capability of the CNN processor in the recognition of characters and human subjects in different environments. The experiments demonstrates impressive results, which proves the proposed digital CNN processor a powerful component in the design of efficient tactile/vision substitution systems for the visually impaired people.**

I. Introduction

THE visually impaired people have confronted inconvenience in their daily lives. Thanks to the recent development in technologies, they now may benefit from the electronic mobility aids that transform visual cues into other sensory modalities.

As early as 1977, Guarniero [1] proposed an electronic device which converted visual information to tactile stimulations. In 1991, Spisz and Weed designed a Tactile/Vision Substitution (TVS) System, which integrated a video camera and a real-time image processing subsystem, was capable of providing a 16×16 vibrator-tactile stimulator array as output [2]. Ram and Sharf verified and confirmed the effectiveness of the TVS systems, and emphasized that the electronic mobility aids could help the visually impaired people in raising their confidence and independence in mobility [3].

In this study, we propose to use Cellular Neural Network (CNN) to implement TVSS**.** CNN was first proposed in 1988 [4][5] and has evolved to cover a wide range of applications. Most of these applications were in image processing [6]-[9]. Using CNN for image processing has advantages in hardware implementation and real-time processing. Several analog CNN chips and universal machines (UMs) based on the chips are now available in the market and their specifications were compared in [10]. However, an efficient approach to study CNN functions is to emulate CNN functions on digital CNN-UM architectures [11]-[13]. This approach utilizes numerical methods to solve the partial differential equations in CNN, such that complex dynamics of the CNN system can be easily implemented and functions be tested.

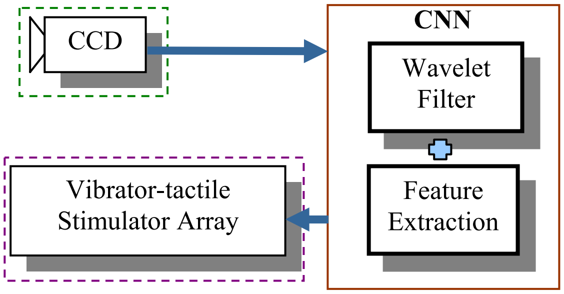

Fig 1. Block diagram of the Tactile/Vision Substitution System (TVSS)

We propose to use CNN as the image processor for a tactile/vision substitution system (TVSS). In a TVSS, a series of images are acquired from the acquisition module. The acquired images must be down-sampled, features be enhanced, and then serve as input to a vibrator-tactile stimulator array. In this study, we focus on the image processing part of the TVSS. The two image processing functions are implemented on a single digital CNN processor with the embedding method published previously from our group [14]. An FPGA board is employed to demonstrate and verify the functions of the digital CNN processor.

II. The Proposed Method

The block diagram of the prospected TVSS is shown in Fig. 1**.** The TVSS contains a CCD camera, a FPGA-based CNN processor, and a vibrator-tactile stimulator array. Since this study focuses on the image processing of the acquired images using CNN, we will briefly introduce the image acquisition module and describe the CNN image processing in details.

A. Image Acquisition:

The image acquisition module includes a CCD camera and an image cutting circuit which limits the output image to a preset size for the following processing. Since the intended output, the vibrator-tactile stimulator array, has 32×32 small vibrators, we need to down-sample the acquired image into that size. However, our preliminary experiments showed that directly using 32×32 images has poor effect on the final results. On the contrary, the information would be better preserved if we acquire larger images and down-sample the

1-4244-0033-3/06/$20.00 ©2006 IEEE.

images into smaller ones. Therefore, we designated to acquire larger , 64×64, images from the original CCD camera and then use the wavelet filter in the image processing to down-sample the images into a size of 32×32 for the following processing.

B. Image Processing:

The image processing subsystem includes a down-sampling circuit and a feature extraction unit, each of which will be described separately in the following:

1. Down-sampling with a wavelet filter

Discrete wavelet transform (DWT) has been used successfully in the analysis of signals and images by decompose them into subband conponents. The 1-D DWT uses well-designed low-pass and high-pass filters to separate the signals into components in different bands. A 2-D DWT can be achieved by applying 1-D DWT to the row elements first, followed by another 1-D DWT in the column elements. The sequential applications of two 1-D DWT operations can be achieved with a single operation of convolving the original image with 2-D wavelet transformation matrices [13]. The matrices are expressed as follows:

$$M_{LL} = \vartheta^T \vartheta, M_{LH} = \vartheta^T \psi, M_{HL} = \psi^T \vartheta, M_{HH} = \psi^T \psi \quad (1)$$

where θ and ψ are the low-pass and high-pass filter operators, respectively, of the 1-D DWT. The resulting images of the four bands are usually down-sampled by 2 both in the row and in the column to produce a decomposed image of the same size.

There are many wavelet filter sets available in the literature. We empirically select DB53 filters in this study. Besides, since our purpose of applying wavelet transformation is to preserve most important visual cues in the down-sampled image, we choose to use the LL band image for the following processing.

2. Features extraction

After the input image has been down-sampled to the desired size, the important visual cues should be outlined with feature extraction operations. In this study, we use a half-toning operator for scene enhancement and feature extraction. The half-toning process is an important technique in image processing that transforms a grayscale image into a binary image that shows an ample sketch of the original grayscale image. CNNs have been shown to produce high-quality half-toning images [15]. We empirically chose the Template 1 in [15], with neighborhood size r =2, for the experiments. Details of this template can be found in [15].

3. The CNN Embedding Method

We have proposed an embedding method that can embed multiple image processing functions into a single CNN processor [14]. The block diagram of this method is depicted in Fig. 2. With this method, sequential image processing functions can be merged into one operation, with negligible differences in the result. In the sequel, the calculation load can be soothed and speed be highly enhanced.

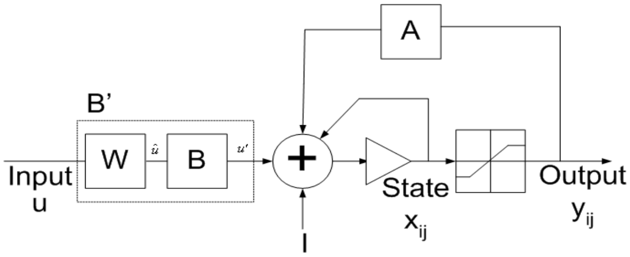

Fig. 2 Block diagram of the CNN template embedding method.

In this study, we intended to embed a 2-D wavelet decomposition filter with a half-toning function into a single digital CNN processor. The CNN half-toning function belongs to the regular CNN functions with the B template not equal to zero (B≠0). In order to combine a regular CNN template B (≠0) with the wavelet filter (W), the new B template (B') is the convolution of B template with W [14], such that:

$$\begin{aligned} u'(i,j) &= \sum_{C(k,l)\in Nr(i,j} B(i,j;k,l)\hat{u}_{kl} \\ &= \sum_{C(k,l)\in Nr(i,j)} B(i,j;k,l) \sum_{C(k,l)\in Nr(i,j)} W(i,j;k,l)u_{kl} \end{aligned} \quad (2)$$

In order to calculate Eq. (2), a CNN with twice the neighborhood size must be used. If we want to use a CNN of the same neighborhood size to implement the resulting B' template, some approximation must be made [14] that

$$\begin{cases} \hat{u}_{kl} = \sum_{C(k,l)\in Nr(i,j)} w(i,j;k,l)u_{kl} & \textit{for } \hat{u}_{kl} \textit{ at k=0 and l=0} \\ \hat{u}_{kl} = u_{kl} & \textit{otherwise} \end{cases} \quad (3)$$

In the sequel, the CNN equation can be modified as

$$\begin{aligned} \frac{dx_{ij}(t)}{dt} = -x_{ij}(t) + \sum_{C(k,l)\in Nr(i,j)} A(i,j;k,l)y_{kl}(t) + \\ \sum_{C(k,l)\in Nr(i,j)} B'(i,j;k,l)u_{kl}(t) + I \end{aligned} \quad (4)$$

and template B' can be derived as

$$B'(i,j) = B(i,j) + B(i,j;0,0) \times [W(i,j) - O(Nr)] \quad (5)$$

where the $O(Nr)$ is a $(2r+1)\times(2r+1)$ matrix with a center element of unity and the other elements are zero.

4. A Numerical Example of the Approximation Methods

The relationship templates of the DB53 wavelet for 1-D wavelet decomposition are

$$\vartheta = \begin{bmatrix} \frac{-1}{8} & \frac{1}{4} & \frac{3}{4} & \frac{1}{4} & \frac{-1}{8} \end{bmatrix}, \quad \psi = \begin{bmatrix} \frac{-1}{2} & 1 & \frac{-1}{2} \end{bmatrix}$$

The 2-D wavelet transformation matrix of DB53 in the M_{LL} band is calculated as the outer product of θ as

$$W = M_{LL} = \begin{bmatrix} \frac{1}{64} & \frac{-1}{32} & \frac{-3}{32} & \frac{-1}{32} & \frac{1}{64} \\ \frac{-1}{32} & \frac{1}{16} & \frac{3}{16} & \frac{1}{16} & \frac{-1}{32} \\ \frac{-3}{32} & \frac{3}{16} & \frac{9}{16} & \frac{3}{16} & \frac{-3}{32} \\ \frac{-1}{32} & \frac{1}{16} & \frac{3}{16} & \frac{1}{16} & \frac{-1}{32} \\ \frac{1}{64} & \frac{-1}{32} & \frac{-3}{32} & \frac{-1}{32} & \frac{1}{64} \end{bmatrix}$$

The CNN half-toning B, template 2 in [15] is

$$B = \begin{bmatrix} 0.02 & 0.07 & 0.10 & 0.07 & 0.02 \\ 0.07 & 0.32 & 0.46 & 0.32 & 0.07 \\ 0.10 & 0.46 & 0.81 & 0.46 & 0.10 \\ 0.7 & 0.32 & 0.46 & 0.32 & 0.07 \\ 0.02 & 0.07 & 0.10 & 0.07 & 0.02 \end{bmatrix}$$

The resulting B' templates can be calculated from Eq. (5) and the result is summarized as follows:

$$B' = \begin{bmatrix} 0.0311 & -0.0158 & -0.0271 & -0.0158 & 0.0311 \\ -0.0158 & 0.2392 & 0.4217 & 0.2392 & -0.0158 \\ -0.0271 & 0.4217 & 0.5625 & 0.4217 & -0.0271 \\ -0.0158 & 0.2392 & 0.4217 & 0.2392 & -0.0158 \\ 0.0311 & -0.0158 & -0.0271 & -0.0158 & 0.0311 \end{bmatrix}$$

5. The Emulated Digital CNN Implementation

The analog array structure of the CNN Universal Machine (CNN-UM) was proposed in 1993 to implement the original analog CNN paradigm [16]. Later, in 1994, the digital version of CNN-UM was published in an attempt to provide an efficient emulator system for the analog CNN [11]. Several different hardware structures and emulation strategies have been proposed to improve the performances of the digital CNN-UM [12][13]. In this study, we intended to employ the distributed coefficient method [11] to implement the Digital CNN image processor for the TVSS.

We used an FPGA board (VIRTEX-II 2000, XILINX) to implement the digital CNN image processor in the TVSS. The FPGA board is an embedding system which uses XILINX xc2v2000 chips to implement the digital circuits and use MicroBlaze^{TM} as the soft processor core of the system. The acquired image data was first transmitted into the local memory on the FPGA board. These data serve as inputs to the digital CNN processor on the FPGA chips. The operations of the system and the control of data flows are managed by the MicroBlaze core through the OPB data bus. The digital CNN processor was designed into an intellectual property (IP), in a form that benefit from the integrality and reusability of the FPGA circuit.

III. EXPERIMENT AND RESULT

Three experiments were designated to test the performance of the system. In the first experiments, we test an important application of the TVSS system: the character recognition. The ability to recognize characters can help the visually impaired people regain some of their ability to read books and acquire useful information, such as road signs, from the environment. The characters were first captured with the CCD camera. The images were manipulated with the digital CNN image processing unit, and then sent to the vibrator array as output. Fig. 3 shows the original images and the processed results of the characters. Fig. 3(a) is a Chinese character "Lin", (b)-(d) are the English letters "Q", "R', and "S", respectively. The character and letters can be easily recognized. The results show that the combination of wavelet filter and the half-toning function is proven to work quite well in the extraction of features for character recognition.

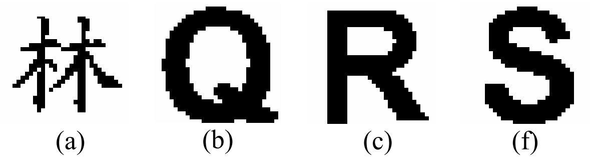

Fig. 3 The processed image of (a) Chinese character "Lin", and English letters (b) "Q", (c) "R", and (d) "S", respectively.

The second experiment is the processing of a sequence of images when s subject is walking across the scene with a white background. The result is depicted in Fig. 4. The images in the upper row show the original images and that in the lower row are the images after CNN wavelet filtering, down-sampling, and half-toning. The subject can be easily located in the processed images, even with the very low resolution of 32×32. The upper part of the subjects which shows quite different gray-level with the background is obvious in the images. The lower part of the subject, which is similar to the background in color, also can be distinguished from the background, although some blurs may be seen.

The third experiment is to test the performance of the system with complex background such as buildings and stairs, which is usually concurred in an outdoor environment. A subject was asked to stand on the stairs while the TVSS was moving slowly toward the subject in order to mimic the mobile experience of the TVSS users. The results are shown in Fig. 5. Similar to Fig. 4, images in the upper row show the acquired images and that in the lower row are the processed results. These images show an impressive result of this system. The subject can be easily located in the processed images. However, some information of the background may be lost, probably because of the down-sampling process and the limited resolution of the intended tactile vibrator array.

Please note that the size of the binary output image is only a quarter of the input image. This is confined by the fact that the high-resolution visual information must be transformed to the much lower resolution of the tactile sensation. Therefore, just as suggested in many previous TVSS reports, an (electric) guide stuck is usually recommended when using the TVSS in the outdoor environment. However, from the impressive results demonstrated in our experiments, the use of CNN as image processor in a TVSS did prove to be a good solution of the high computation problem of the system. The application

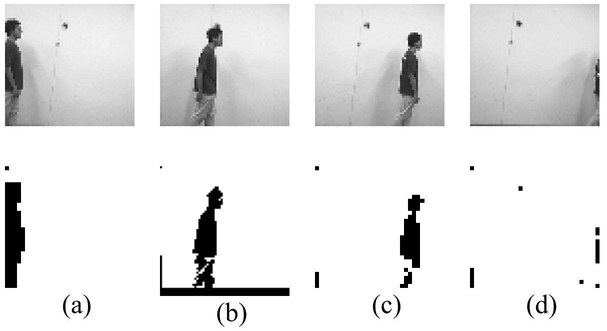

Fig. 4 Experimental results when a subject is walking across a white background, from (a) left-hand side, to (b) middle, to (c) right-hand side, and (d) out of the scene. Upper images are the images acquired by the CCD camera. Lower images are the processed results.

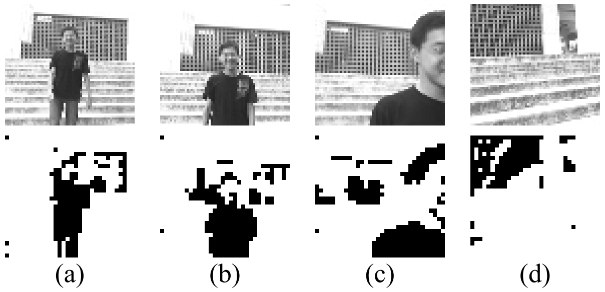

Fig. 5 Experimental results when the TVSS user is walking (a)-(c) toward a subject and (d) passing by the subject in an outdoor environment with complex background. Upper images are the images acquired by the CCD camera. Lower images are the processed results.

of the embedding methods to combine multiple CNN functions into a single CNN processor further enhances the efficiency of the system.

IV. CONCLUSION

In this paper, we proposed to use a digital CNN processor to solve the image processing problems in a TVSS. The CNN processor was used to perform 2-D wavelet filtering (and down-sampling) and half-toning functions. The template combination method was exploited to merge the two functions into a single CNN operation. This embedding method simplifies the system architecture and decreases the size of the circuit, which, consecutively, helps to increase the operation speed and decrease the power consumption of the system. The digital CNN image processor was test in character recognition and in the recognition of human subjects in both simple and complex environments. The experiments demonstrated satisfactory results.

In the future works, we designate to combine this CNN processor with a vibrator-tactile stimulator array to test the capability of the entire TVSS. We hope that the visually impaired people will find this system efficient and beneficial to their lives.

ACHNOWLEDGMENTS

This study was supported in part by the grants NSC91-2614-E-194-001 and NSC92-2218-E-194-002 from the National Science Council, Taiwan, R.O.C.

REFERENCES

[1] G. Guarniero, “Tactile vision a personal view.” Viz. Impairment and blindness, pp.1225-130, Mar. 1977

[2] T. S. Spisz and H. R. Weed, “An image acquisition subsystem for tactile vision substitution”, Proc. Ann. Int’l Conf. IEEE EMBS, 1991, Vol. 13, pp.1835-1836, Oct.31-Nov. 3 1991.

[3] S. Ram and J. Sharf, ”The People Sensor: A mobility aid for the visually impaired”, Wearable Computers pp. 166-167, Oct. 1998

[4] L. O. Chua and L.Yang, “Cellular neural networks: Theory,” IEEE Trans. Circuits Syst., Vol. 35, pp. 1257-1272, 1988.

[5] L.O. Chua, and L.Yang, “Cellular neural networks: Applications,” IEEE Trans. Circuits Syst., Vol. 35, pp. 1273-1290, 1988.

[6] P. Foldesy, L. Kek, T. Roska, A. Zarandy, and G. Bartfai, “Fault-tolerant design of analogic CNN templates and algorithms-Part I: The binary output case”, IEEE Trans. Circuits Syst., Vol.46, no.2 pp.312-322, Feb. 1999.

[7] T. Roska, “Cellular wave computers for brain-like spatial-temporal sensory computing”, IEEE Trans. Circuits Syst., Vol.5, no.2, pp.5-19, 2005.

[8] C.-T. Lin; C.-L. Chang; J.-F. Chung, “New horizon for CNN: with fuzzy paradigms for multimedia”, IEEE Trans. Circuits Syst., Vol.5, no.2, pp.20-35, 2005.

[9] T.-J. Su, Y.-Y. Du, Y.-J. Cheng, and Y.-H. Su “A fingerprint recognition system using cellular neural networks”, Proc. IEEE Int’l. Workshop Cellular Neural Networks Applications, CNNA 2005, pp.170-173, May 2005.

[10] T. Roska and Á. Rodriguez-Vázquez, “Towards Visual Microprocessors”, Proc. IEEE, Vol. 90, no.7, pp.1244-1257, Jul. 2002.

[11] K. A. Wen, J. Y. Su and C. Y. Lu, “VLSI design of digital Cellular neural networks for image processing”, Journal of Visual Communication and Image Representation, Vol.5, No.2, pp.117-126, Jun. 1994.

[12] Z. Nagy, P. Szolgay, “Configurable Multilayer CNN-UM Emulator on FPGA”, IEEE Trans. Circuits Syst. I: Fundamental Theory and Applications, Vol. 50, No.6 pp.774-778, Jun. 2003.

[13] Z. Nagy, P. Szolgay, “Numerical solution of a class of PDEs by using emulated digital CNN-UM on FPGAs”, Proc. 16th European Conf. on Circuits Theory and Design, Cracow, Sep. 2003.

[14] C.-N. Lin, S.-N. Yu, W.-C. Chuang, “Embedding of Multiple Functions into a Single Cellular Neural Network an Image Filtering Prospective”, Proc. IEEE Int. Workshop Cellular Neural Networks Applications, CNNA 2005, pp.166-169, May 2005.

[15] K. R. Crounse, T. Roska, and L.O.Chua, “Image halftoning with cellular neural networks”, IEEE Trans. Circuits Syst., Vol. 40, pp. 267-283, Apr. 1993.

[16] T. Roska and L.O. Chua, “The CNN Universal Machine: An analogic array computer”, IEEE Trans. Circuits Syst. II, Vol. 40, pp. 163-173, Mar. 1993.