Extended Abstracts of the 2006 International Conference on Solid State Devices and Materials, Yokohama, 2006, pp. 588-589

P-5-3

A Novel Vision Chip for High-Speed Target Tracking

Wei Miao, Qingyu Lin and Nanjian Wu

Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, P. R. China
Phone/Fax: +86-10-8230-4754 E-mail: nanjian@red.semi.ac.cn

1. Introduction

One of challenges in the field of image processing is to develop a target tracking vision chip. Some high-speed target tracking vision chips have been reported by several groups. An analog vision chip can focus attention on a salient target was developed [1]. Another target tracking vision chip was realized by foveated silicon retina [2]. Digital vision chips for target tracking based on self-windowing method were also developed [3, 4]. These chips can perform real-time image processing at the rate of 1000fps. However, they only implement special algorithms so that they could only be applied in some ideal cases.

This paper proposes a target tracking vision chip based on novel algorithms. We developed two concise algorithms that can be used to process complex target images at a high speed. The algorithms consist of efficient operations: such as collision detection, separation detection and position extraction. The operations cost less time and resources than those presented in [3, 4]. Further, the chip uses row-parallel processing architecture to save chip area. The chip can perform target tracking at the rate of 1000fps with more functionality and less area than the reported digital chips.

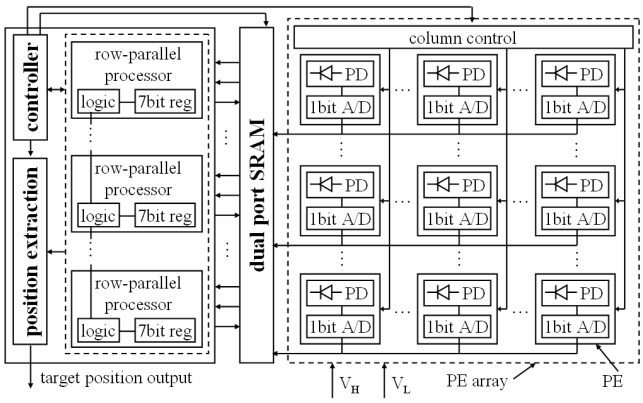

Fig. 1 Architecture of the novel target tracking chip.

2. Chip architecture

Fig. 1 shows the architecture of the proposed high-speed target tracking chip. It consists of a pixel element (PE) array, a dual port SRAM, row-parallel processors, a position extraction module and a controller module.The PE array obtains N×N-pixel image and binaries it. Each PE has a photodiode (PD) and a 1bit A/D, shown in Fig. 2. A PD_ctrl signal controls the integration time of the PD optical signal. A PE outputs logic 1 if the PD reverse voltage is between two threshold voltages V_L and V_H ($V_L<V_H$), else it outputs logic 0. The two thresholds can make the target clearer and reduce non-target objects in binarized images. The PE array row-parallelly outputs digital data to the dual port SRAM with 4N×N bits. The row-parallel processors with 7-bit registers can read two columns of data from SRAM during each read operation, and can perform fifteen kinds of operations. The position extraction module gets the position of the target and sends it out of the chip. Following the algorithms, The controller module controls image acquisition and image binarizing in the PE array, image processing in the row-parallel processors, and target position extraction in the position extraction module. The chip works at a rate of 1000fps.

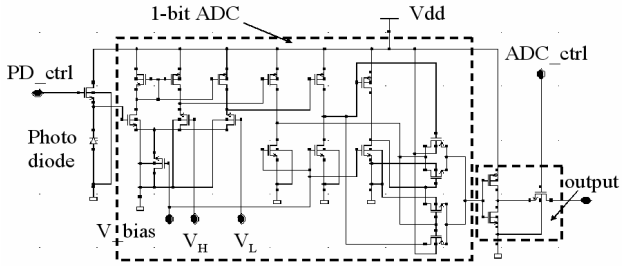

Fig. 2 The circuits in a PE.

3. Algorithm

Two algorithms are implemented in the chip. Algorithm 1 is suitable to track a high-speed target in relative static background. Algorithm 2 is suitable to track a high-speed target among other high-speed objects in clear background. Fig. 3 is the flow chart of the two algorithms.

Let A(x,y) represent a binary image. In void space, A(x,y)=0. In objects, A(x,y)=1. We define some functions:

dilate(A):

dilate(A)(x,y)=A(x+1,y)|A(x-1,y)|A(x,y+1)|A(x,y-1)|A(x,y)

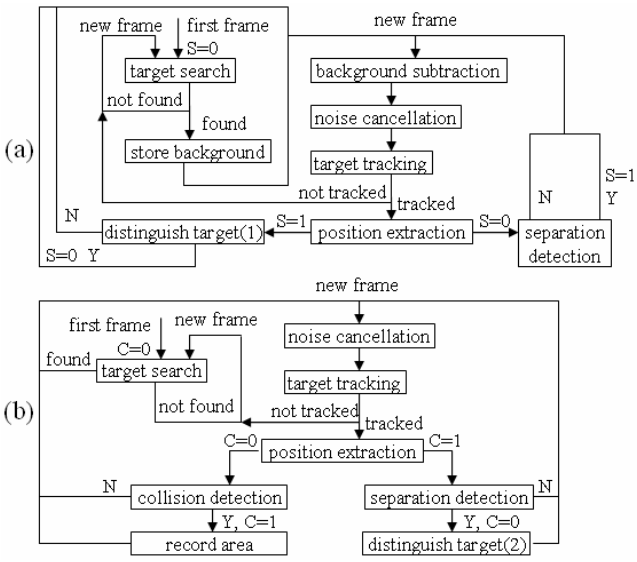

Fig. 3 Flow chart of algorithm 1 (a) and algorithm 2 (b).

Apo(A): (shown in Fig. 4(a))

Apo(A)='OR' of all the pixels' value A(x,y) in the image A

RG(A,B): (time cost: denoted as t_{rg})

This function performs region growing. There is a seed in image A and a region in image B. If the position of the seed is in the range of the region, the seed will grow referring to image B and form a equal region as that in image B. Fig. 4(b) shows the region growing process in row-parallel way.

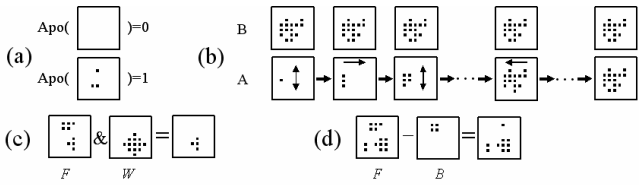

Fig. 4 Some operations: (a) Function Apo. (b) Row-parallel region growing. (c) Target search. (d) Background subtraction.

Denote target image as T(x,y), the image from sensors as F(x,y), and the stored background image as B(x,y). Some operations in the algorithm flow are explained below.

Target search: (shown in Fig. 4(c))

Set a window W(x,y) at certain position with certain radius to find the target. At the first frame, the window position is the center of the image. When the target is lost, the window position is the last position of the target. If the target is found, Apo(W&F)=1; else, Apo(W&F)=0.

Background subtraction: (shown in Fig. 4(d))

F(x,y)=~B(x,y)&F(x,y) or F=F-B;

Target tracking:

Use the self-windowing algorithm[3], but repeat it for m (e.g. m=2) times to cover random distortion of the target.

Repeat m times: T(x,y)=F(x,y)&Dilate(T)(x,y)

Collision detection: (time cost: 4N+1 cycles)

Collided: Apo(F(x,y)&Dilate(T)(x,y)-T(x,y))=1;

Not collided: Apo(F(x,y)&Dilate(T)(x,y)-T(x,y))=0.

Separation detection: (time cost: $<t_{rg}+4N+1$ cycles)

1. Let S(a,b)=T(a,b)=1; S(x,y)=0, x≠a, y≠b.
2. Separated: Apo(T(x,y)-RG(S,T)(x,y))=1;
   Not separated: Apo(T(x,y)-RG(S,T)(x,y))=0.

One part of the separated target is obtained at the time that separation is detected.

Position extraction: (time cost: <2N cycles)

Let x_{min} and x_{max} denote the left edge and the right edge of the target separately, and let y_{min} and y_{max} denote the bottom edge and the top edge of the target separately. The target position is its center:

$$\{x_c, y_c\}=\{(x_{min}+x_{max})/2, (y_{min}+y_{max})/2\}$$

Comparing with the centroid extraction operation, the center extraction operation is easier, faster and sufficient.

Distinguish target (1): (time cost: <4N cycles)

The target is affirmed as the one of the separate parts, which moved more than s (e.g. s=5) pixels

Distinguish target (2): (time cost: 5N+3 cycles)

In the two separate parts, the target is affirmed as the one that has less difference between its area and the area of the target before collision.

The time t_{rg} of region growing depends on the scale and shape of the grown region. Typically, it is propotional to N^{α}, (α is near one and large than one). The worst case is $3N^2/4$, which would not happen in real application. The maximum processing time cost in one frame is $19N+t_{rg}$. So chip main clock of 10MHz is enough for N=64, and clock of 50MHz is enough for N=256 to satisfy the rate of 1000fps.

4. Chip implementation and experiments

A prototype target tracking vision chip with 64x64 PE array is fabricated using 0.35 μm standard CMOS process. Fig. 5 is the photo of the chip and one frame snapshot. Table I gives its specifics. We first adjusted V_H and V_L to set a window just captures the lightness of the target, and then start one of the two tracking algorithms. A high-speed target is tracked by algorithm 1. The target itself has fast variation in shape and the background objects have slow variation. The target track is shown in Fig. 6(a). One of two high-speed objects that have distinct difference in lightness with the background is tracked. The two objects collided and then separated. During the period, their shapes vary little. This is shown in Fig. 6(b).

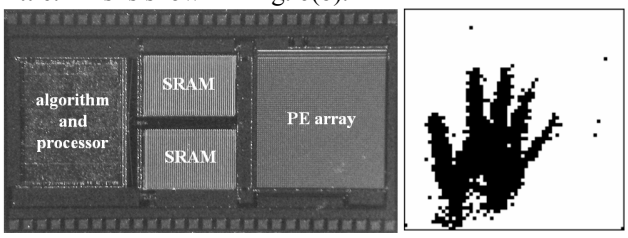

Fig. 5 Photo of the test chip and one frame snapshot.

Table I specifics of the proposed chip

process	Chartered 0.35μm CMOS
resolution	64×64
chip area	4.5mm×2.5mm
pixel area	20μm ×20μm
photo diode area	6μm ×6μm (fill factor 9%)
frequency	10 MHz
power consumption	30mW

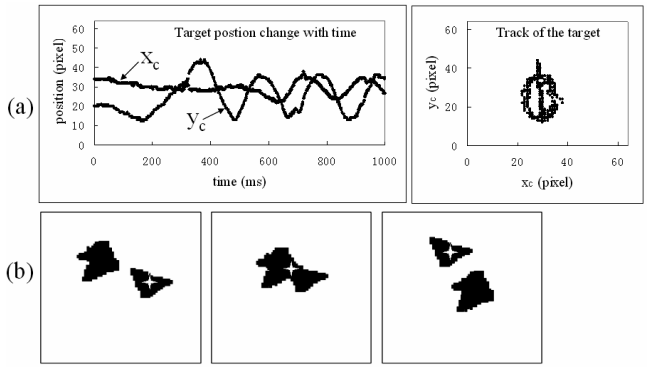

Fig. 6 (a) the track of the target in algorithm 1. (2) algorithm 2.

References

[1] V. Brajovic, and T. Kanade, IEEE J. Solid State Circuits, **33** (1998) 1199.

[2] R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, and M. Zhang, IEEE Trans. Circuits Syst. II, **47** (2000) 504.

[3] I. Ishii, Y. Nakabo, and M. Ishikawa, *Proc. IEEE Int. Conf. on Robotics and Automation* (1996) 2309.

[4] T. Komuro, I. Ishii, M. Ishikawa, and A. Yoshida, IEEE Trans. Electron Devices, **50** (2003) 191.