An Object Tracking Computational Sensor

Vladimir Brajovic

CMU-RI-TR-01-40

Robotics Institute, Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh PA 15213
ph: (412)268-5622, fax: (412)268-5571
brajovic@ri.cmu.edu

December 2001

Abstract

An analog VLSI computational vision sensor detects an optical image of multiple intensity–coded objects. As the image is being sensed, the sensor computes and reports the area and the x,y location of the object on three analog pins. A 43x43 cell prototype is implemented in 2µ CMOS technology. Each cell/pixel is 47x47µm with a 30% fill factor.

1. Introduction

The paper describes an analog VLSI computational vision sensor for computing the area and centroid of multiple intensity-coded objects. The area and the x,y location of the objects are read directly on three analog pins, as the image is being sensed. The low latency operation makes it suitable for time-critical applications such as head/eye tracking for human machine interfaces. A 43x43 cell prototype is implemented in 2u CMOS technology. Each cell/pixel is 47u x 47u with a 30% fill factor.

Traditionally, position sensitive detectors (PSD) compute the position of an object as the centroid of the object's spatial intensity distribution. Since the object is not segmented from the background, the traditional approach works only for a well defined small bright object (e.g., a spot) against the dark background. Standley [4] implemented segmentation of the object from the background by subtracting a user-defined threshold current from a photocurrent in each pixel. This nulls the influence of the background. As the devices must operate in the subthreshold region, the threshold currents significantly mismatch among pixels; the image of the object is thus corrupted: the accuracy of the chip is acceptable only for large objects which average over the spatial noise. In principle, the Standley's chip can (cumulatively) detect multiple intensity–coded objects by sweeping the threshold level. Due to the threshold current mismatch, however, detecting only very bright objects is practical, whereas darker objects are dominated by the introduced noise. Standley's pixel takes approximately 190u x 190u.

2. Implementation

In our implementation, an optical image of the scene is focused and sensed directly on the chip. Using the intensity–to–time processing paradigm [1], the chip continuously segments the bright objects from the dark background by evolving a binary image. In the intensity–to–time paradigm, the pixels receiving stronger stimuli respond before the pixels receiving weaker ones. Early on after the frame reset, no pixels are turned on, and the entire binary image is zero. After some time, the brightest pixels respond, forming a non-zero object in the binary image. Some time later, intermediate intensity pixels respond, thus adding and evolving new objects. Eventually all the pixels respond; the entire image becomes non-zero. From such an evolving binary image, the chip continuously computes the area and centroid of the distribution of non–zero pixels, using the method of projections [3] and two 1D resistive grids [2].

The block diagram of the chip is shown in Figure 1. The pixel circuit is shown in Figure 2. All cells are reset simultaneously, after which they begin to integrate the photo–charge. When a predetermined amount of photo–charge is collected within the sensing element, the inverter trips and turns on two constant current sources M1 and M2. The current sources represent the binary image of the input scene. Each of the two current sources is summed into a vertical and horizontal bus, respectively. This is the method of projections which decomposes the 2D problem into two 1D current distributions. The total current I_T on the horizontal (or vertical) bus represents the total number of cells that are turned on, i.e., the area of the object. The centroids of horizontal and vertical 1D distributions represent the x and y object centroid, respectively. These quantities are found from peripheral currents I_x and I_y of the horizontal and vertical resistive grids as:

$$\bar{x} = (N-1)\frac{I_x}{I_T} + 1$$

$$\bar{y} = (M-1)\frac{I_y}{I_T} + 1$$

where I_T is a total current in the distribution (i.e., the object area), M and N are the total number of rows and columns, respectively. In our chip $M=N=43$. The resistive grids are made of the polysilicon strip.

To avoid division by I_T, the currents from the vertical and horizontal busses are fed into a normalizing current mirror [2]. The normalizing current mirror guarantees that the current distribution flowing out of the mirror always sums up to a user–defined *constant* current I_S. Thus, the currents flowing from the periphery of the resistive grid is directly proportional to centroid x and y irrespective of the I_T.

3. Experiments

Typical test patterns used in an early experimentation consist of overlapping (Figure 3a) and non overlapping (Figure 4a) discs of varying intensity. The waveforms for the area and the (x,y) position are monitored by a digital scope. For each test pattern, the evolving (x,y) position of the centroid is viewed in the scope's XY mode. The temporal trajectories of the centroid are shown in Figure 3b and Figure 4b. The intensity of the trace encodes the duration spent at a particular location. We observe that the centroid transitions among three stable locations. From the geometry of the pattern and the optical setup, we overlay three circles in Figure 3b and Figure 4b. Some error can be observed, but general behavior is as expected. A more rigorous error analysis is currently underway, but the early hypothesis suggests that the light impinging on the array modulates current sources M1 and M2. Namely, the cells within bright objects will contribute more current than the cells from darker objects. Therefore, the bright objects appear “heavier” than they actually are. For example, the position of the darkest disk in Figure 3 is biased toward the encapsulated bright small object.

Figure 5 shows typical waveforms for the three disk examples. The top trace represents the total object area. The area monotonically increases as more and more pixels are turned on. The waveform of the area also represents the scene's cumulative histogram: a temporal derivative of this waveform gives a conventional histogram of the scene [1].

We can observe four sharp rising edges in the object area waveform: the three objects and the background. After each of the objects is turned on, the waveforms for the (x,y) position are updated and are stable for sampling.

4. Conclusion

A vision computational sensor chip for multiple object tracking has been built. Preliminary experiments confirm that it is fairly accurate for objects of varying sizes.

[1] V. Brajovic and T. Kanade, “A Sorting Image Sensor: An Example of Massively Parallel Intensity-to-Time Processing for Low-Latency Computational Sensors,” Proc. of the 1996 IEEE Intl. Conf. on Robotics and Automation.

[2] S.P. DeWeerth, “Analog VLSI Circuits for Stimulus Localization and Centroid Computation,” *Intl. Jour. of Comp. Vision*, Vol. 8, No. 3, 1992, pp. 191-202.

[3] B. Horn, *Robot Vision*, MIT Press, 1986.

[4] D.L. Standley, “An Object Position and Orientation IC with Embedded Imager,” *IEEE Jour. of Solid–State Circuits*, Vol. 26, No. 12, December 1991

.

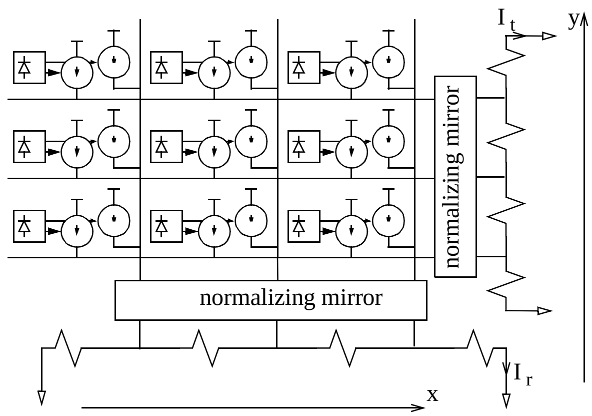


Figure 1: The sensors block diagram.

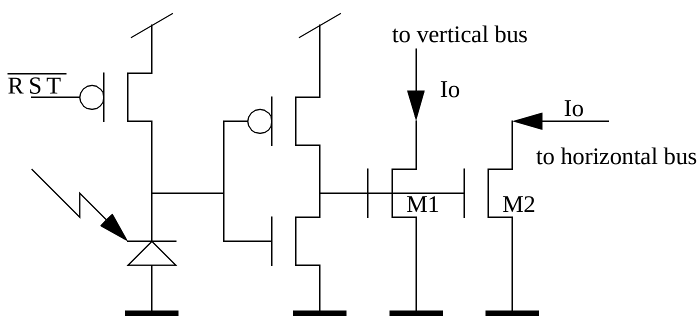
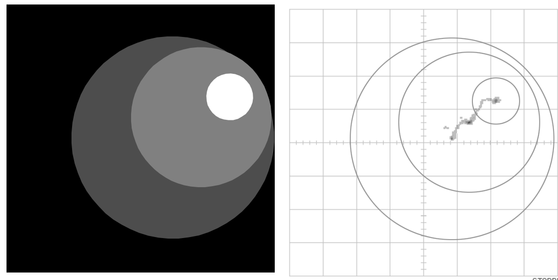
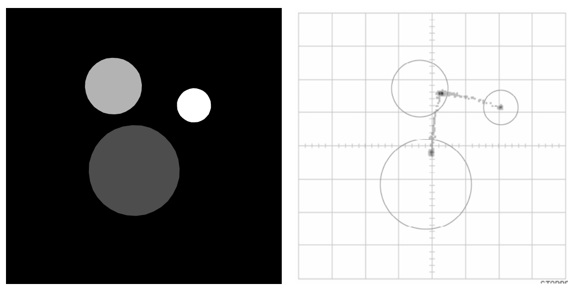


Figure 2: The pixel/cell diagram.



a) b)

Figure 3: Experiment 1: a) test pattern, b) evolution of the centroid.



a) b)

Figure 4: Experiment 2: a) test pattern, b) evolution of the centroid.

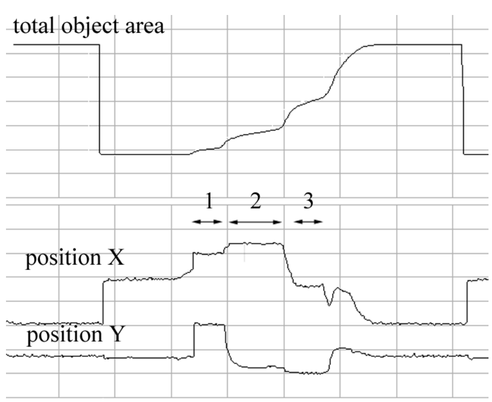


Figure 5: Typical waveforms for the test patterns displayed in Figure 3 and Figure 4.