Proceedings of the 2004 IEEE
International Conference on Robotics & Automation
New Orleans, LA • April 2004

Fast Row-Parallel CMOS Range Image Sensor

Vladimir Brajovic
The Robotics Institute, Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, USA
brajovic@cs.cmu.edu

Abstract—**The paper presents a novel row-parallel architecture for a fast, high-resolution CMOS range sensor utilizing dynamic triangulation. The sensor detects the location of the structured light (e.g., laser line) in each row based on temporal signatures from the sensory signal. The range map is detected at twice the resolution than the photodetector count and is sampled on a 30 x 30 μm grid in the image plane. Our current prototype can operate up to 2000 frames per second, and deliver sub millimeter precision from a stand off distance of about 70 cm.**

Keywords—3D imaging sensors; computational sensors.

I. INTRODUCTION

Sensing range and object shapes is important in many applications and has been an active area for several decades. Yet, today there are no accurate and inexpensive ranging image sensors that would stream range maps at video speed. In this paper we present a CMOS smart image sensor that uses dynamic triangulation to achieve real-time range map streaming.

A traditional structured light triangulation has been found to be most practical for robotic applications [1]. The triangulation setup and geometry is shown in Figure 1. A sensor, usually a CCD camera, views the scene. A stripe of laser light is projected onto the scene from an oblique angle creating a light contour on the object that is descriptive of the local object shape. Once the location of this contour is found in each row (x_o), a slice of range image is calculated from the triangulation geometry. This process is too slow since each slice requires at least one camera frame time. There are well-known alternatives to this method in which a structured light with several stripes is projected. This approach somewhat reduces the number of CCD frames required. However, the burden of transferring, storing and processing several full CCD frames remains, rendering the whole process slow.

A high-speed triangulation, also called dynamic triangulation, continuously sweeps the laser in Figure 1 across the scene, for example, from right to left [2]. Each pixel in the sensor has its own line of sight and "sees" the laser stripe only once as it sweeps by. By recording the time to when a particular pixel at location x_o sees the laser (in respect to a frame trigger/reference), the triangulation geometry is

This work has been supported by the National Science Foundation under Grant No. 0102272.

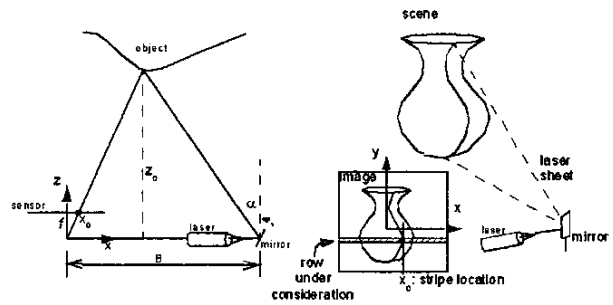

Figure 1: Triangulation ranging setup involving laser structured light.

determined for that pixel and the range along the line of site for that pixel can be calculated. Specifically, the depth to the point on the object is found as:

$$z = \frac{B}{\frac{x_o}{f} + \tan\alpha_o} = \frac{B}{\frac{x_o}{f} + \tan\omega t_o} \qquad (1)$$

where B is the baseline separation of the laser and the camera optical centers, f is the focal length of the camera lens, x_o is the image location within a row where the laser stripe is detected. In dynamic triangulation, the laser projection angle α_o is found by measuring the time when the location x_o sees the laser, i.e., $\alpha_o = \omega t_o$. Therefore, a dynamic triangulation sensor captures range images by collecting (x, t) pairs within each row.

The above-described technique has been implemented in two cell–parallel VLSI range sensors [3][4]. In [3] each cell detects the temporal intensity peak; at that time, the cell records its time–stamp. In [4], each cell includes two photodetectors; the stripe is detected when the appreciable difference between the two photocurrents is observed; then a timestamp is associated with each cell. Even though these VLSI implementations promise high speed and reduced system complexity, the accuracy and reliability are questionable. Most notably the design in [3] stores the timestamp in analog domain; the data degrade over time, resulting in decreased accuracy. On the other hand, a cell in [4] spatially differentiates the signal; therefore, it is "noisy" and is prone to trigger before seeing the laser stripe.

The above-described sensors [3][4] are cell–parallel, i.e., each pixel is concerned with its own illumination. The triangulation method, however, is inherently row–parallel; there is only one pixel in each row that sees the laser stripe at any given time. Therefore, the detection of the image of the laser stripe is a global operation over each row of pixels. This fact allows us to improve the robustness by introducing a broader image context along each row. It also permits that most of the circuitry in [3] and [4] can be removed from each cell and used once–per–row on the side of the array. The pixel area can be improved 50-100x, from about 300x300μm to less than 30x30μm.

In addition to the speed and spatial resolution, the processing at the sensory level in this sensor leverages temporal signatures in the photodetector signal to effectively exclude the influence of ambient illumination and the mismatch of individual photodetectors. As will be shown in a moment, each pixel of the sensor is "tuned" to respond to spikes in the photodetector signal effectively excluding offsets that may be due to ambient illumination, different object albedos and pixel electronics mismatch.

In the remainder of the paper we present our row-parallel architecture for the dynamic triangulation-based range sensor. We also present key circuit details and some experimental results.

II. The Sensor Architecture

Our row–parallel architecture employs a winner–take–all (WTA) circuit embedded in each row of the sensor. The WTA in each row detects the location of the strongest or "most salient" input. In a moment we will show that we use temporal signatures in the signal to encode saliency of the laser; for now, however, assume that the saliency of a pixel is determined by its brightness – the most salient pixel is the brightest one. In addition to the location of the winner, one particular WTA implementation [5] also reports the magnitude of the winning pixel. The location of the winning pixel and its instantaneous intensity is the global row information that is reported as the output for each row.

Figure 2 shows these two signals as a bright laser line moves across the row of 20 photodetectors. A particular cell remains a winner as long as the main portions of the bright target are focused on it. The location of the winning pixel, therefore, appears as a staircase line (Figure 2, top graph). As the stripe is moved, its image leaves one photodetector and begins contributing photocurrent to the next one. At some point, the pixel receiving the target wins and takes control of the common wire reporting the winner's intensity (Figure 2, bottom graph). As the stripe moves toward the center of the new winning cell, the intensity of the winning input current increases. The cell continues to win as the stripe passes the center, but its input current begins to diminish. In the meantime, the next cell begins to receive an increasing amount of light and the process continues. Therefore, as the target

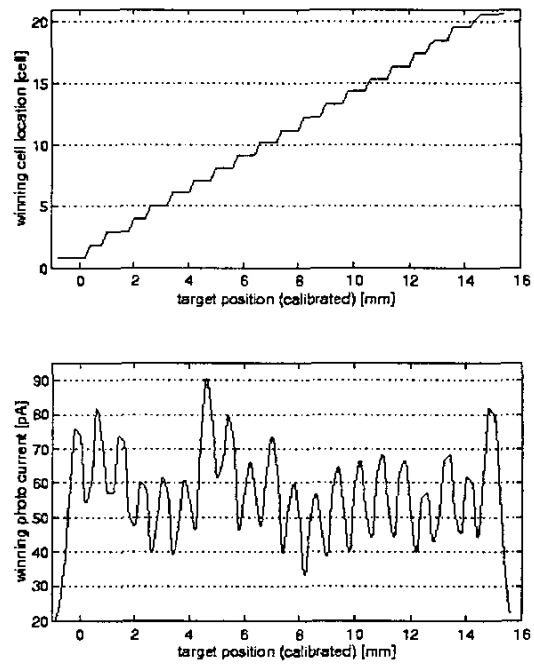

Figure 2: Winning/Losing behavior of Lazzaro's WTA [5]. When the target moves across the row of sensors, the WTA circuit reports pixel location and the intensity of the winner. The peak in the intensity waveform occurs when the target is centered on a photodetector. The valleys occur when the target transitions from one photodetector to the next.

passes over the winning cell, the intensity of the winning input reported on the common voltage [5] increases, peaks, and then decreases. This behavior is clearly displayed in the bottom graph of Figure 2.

The shrinkage of a pixel size due to displacing some of the resource outside of the pixels is only one aspect of improving the spatial resolution. Another aspect of resolution improvement comes from an observation that the positive peaks in the bottom graph of Figure 2 occur when the target is centered on a photodetector. Conversely, the negative peaks (i.e., valleys) occur when the target is positioned exactly between two photodetectors. Therefore, locating peaks (positive and negative) allows precise localization of the laser line at the spatial resolution that is twice the spatial resolution of the photodetectors.

The row-parallel architecture of the sensor is shown in Figure 3. The WTA circuits in each row continuously localizes the laser stripe and generates on its common wire a temporal voltage waveform similar to one shown in the bottom graph of Figure 2. A one-per-row peak detector monitors this voltage. When the peak detector detects either a positive or negative peak in a particular row, it latches the time in one-per-row memory. The row memory locations are rapidly scanned and read out. All rows are scanned several times within the time it takes the laser to travel across one

photodetector. In this way, we ensure that no new peaks occur before we read out information regarding the previous peaks.

When the scanner selects a row, the WTA is multiplexed to the position encoder. At the same time, the type of the peak (i.e., positive or negative) is also multiplexed to the output. Note that the address and the peak type uniquely determine the exact position of the stripe in the image coordinates, that is, x_o in Equation (1). If the peak is positive, the stripe position is in the center of the photodetector whose address is being reported. If the peak is negative, the stripe position is on the receiving edge of the photodetector whose address is being reported. Even though the cell address is not well defined when the stripe transitions to a new cell, the propagation delay through the multiplexer allows the stripe to move into the receiving cell, thus ensuring a stable address. During the readout, the scanner reports the address of the selected row. At the beginning of each frame, a timer is reset. Time information, together with the address and the peak type, provides all the information needed for the reconstruction of the range map according to Equation (1).

III. Selecting Signal Signatures for Encoding Laser Line Saliency

For WTA to select the laser illuminated pixel, the signal from that pixel must be made most salient. If the signal magnitude is used (e.g., trivial saliency), the results may not be satisfactory. The ambient illumination condition and reflectance of object surfaces may not guarantee that the feature created by the structured line contour is the brightest feature along the row of photodetectors. For example, one can imagine that, if the structured light bounces from dark parts of an object, its brightness will easily be less than the brightness produced by the ambient illumination bouncing off of the bright parts of the object. Therefore, a naive algorithm that looks only for brightest features in each row will fail in arbitrary ambient illumination and for an arbitrary object reflectance.

Dynamic triangulation provides a unique opportunity to exploit temporal signatures in the sensory signal to form a powerful saliency for the laser detection. This is because

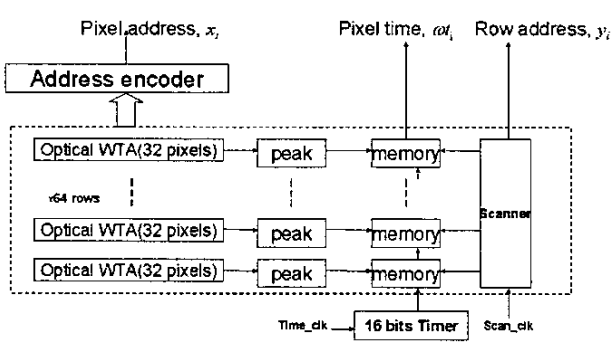

Figure 3: The row-parallel architecture of our range sensor.

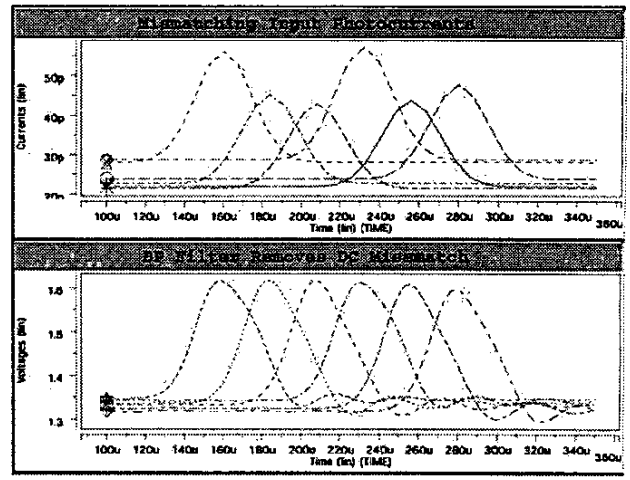

Figure 4: Simulated waveforms for BP pixels: input currents differ among pixels (top graph), BP filter removes this difference (bottom graph).

structured light sweeps across the scene creating sensory signal changes that are much faster than changes induced by ambient illumination or motion of the object reflectance map. These temporal signatures can be extracted by incorporating a band-pass (BP) filter in each pixel. Intuitively, since the BP filter filters out the DC component, the mismatch among pixels becomes insignificant. Additionally, since the BP filter filters out low-frequency components, the effects due to the ambient illumination and object reflectance map are also not significant.

Figure 4 shows simulated waveforms for the performance of the BP pixels feeding the WTA. The top graph shows a photocurrent detected by a number of consecutive photodetectors over which the laser sweeps. Due to the pixel mismatch, object albedo and ambient illumination distribution, we see that these signals have varying amplitudes and offsets. The bottom graph shows the output of the BP filter for each of these pixels. We observe that the unwanted DC component is removed. The filtered signals lend themselves to a more accurate comparison in the WTA circuit.

Figure 5 shows a circuit implementation of the in-pixel BP filter. It is based on Liu's modification [7] of Delbrück's adaptive photoreceptor [6]. Photodetector PD produces a photocurrent proportional to the photo flux impinging on the sensitive area. The feedback circuit comprising of a cascode inverter (M1-M4) establishes the necessary gate voltage on M5 so that M5 can absorb the produced photocurrent. The output of the feedback inverter is taken as the output of the pixel. When the sweeping laser produces a temporal "bump" in the photocurrent, the feedback mechanism attempts to raise the gate of M5 accordingly. Because of the voltage divider comprised of capacitors C_c and C_f ($C_f < C_c$), the pixel output produces a large "bump" just to create a small adjustment on the gate of M5 necessary for M5 to absorb the photocurrent

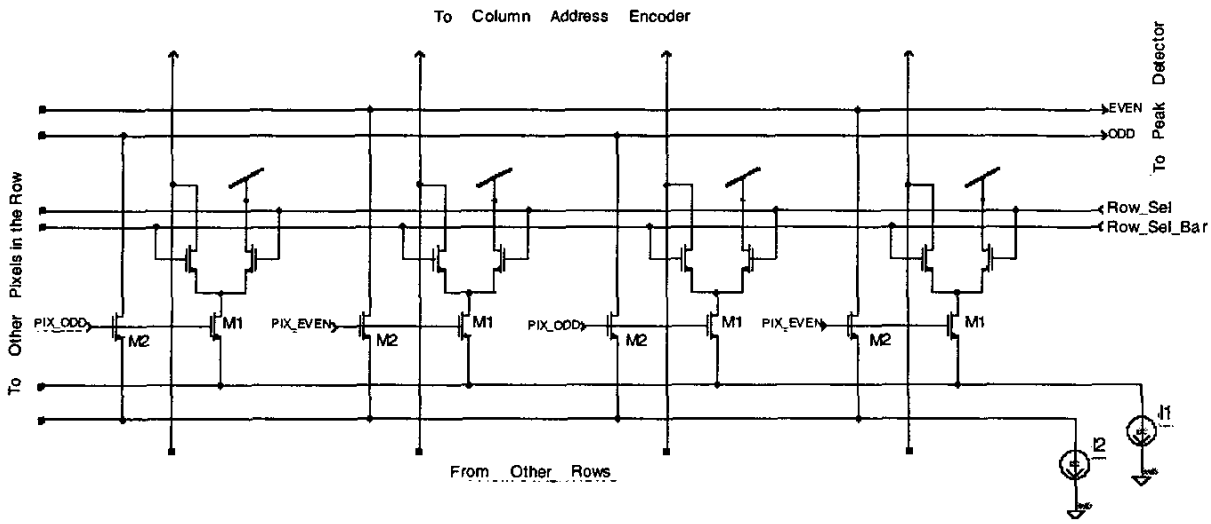

Figure 6: In-row WTA circuit for producing 1) the winner for encoding in the address encoder, and 2) differential current for detection of the peaks in the one-per-row peak detector.

"bump". The DC component is removed by the presence of PMOS M6 that acts as a large tunable resistor. Over a long period of time, the resistor charges the C_f and C_c to a quiescent voltage necessary to absorb the DC photocurrent in M5. In the short run, however, the resistor M5 is effectively an open circuit allowing the pixel output to swing around that quiescent voltage with the gain controlled by the ratio of C_f/C_c. If extremely small energy is returned to the sensor because perhaps the object did not reflect enough of the laser radiation, the sensor will simply report no range measurement for such a condition.

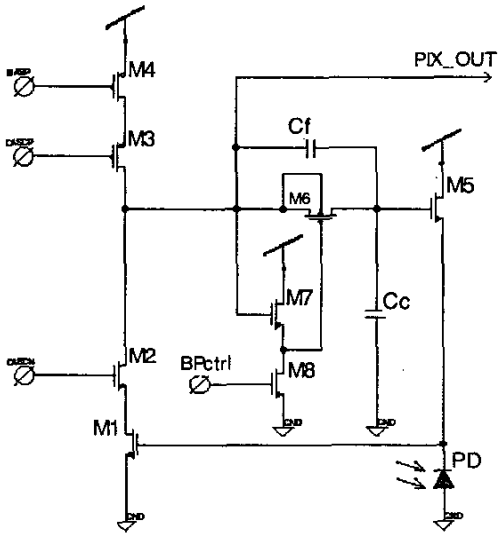

Figure 5: In-pixel band-pass filter (after [7][6]).

IV. The Row Processor Circuit

To make the detection of the peaks in each row more robust we implemented the row processor shown in Figure 6. The figure shows four pixels, two at odd positions and two at even positions. Each pixel receives a voltage signal from its own in-pixel BP filter (Figure 5). Source coupled transistors M_1 together with the current source I_1 comprise current-mode WTA. The largest pixel signal will take the largest portion of the current I_1. The output currents are sent to the peripheral address encoder when the row scanner selects the row.

For the precise localization of the pixel, another set of source-coupled transistors (M_2 and I_2) is included. Currents from odd M_2 transistors are summed in an "odd" wire, and currents from even M_2 transistors is summed up in an "even" wire. The odd/even wire therefore, contains a differential

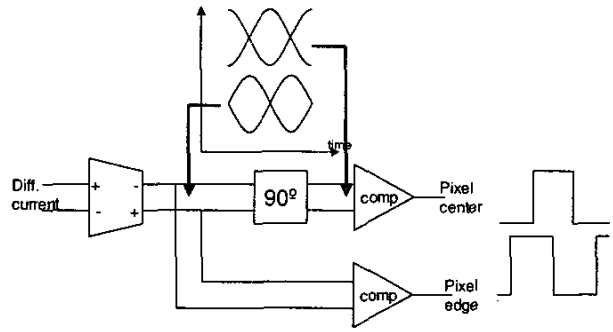

Figure 7: Block diagram of the peak detection circuit for localization of the laser at 1) the centers of photodetectors and 2) the edges between photodetectors.

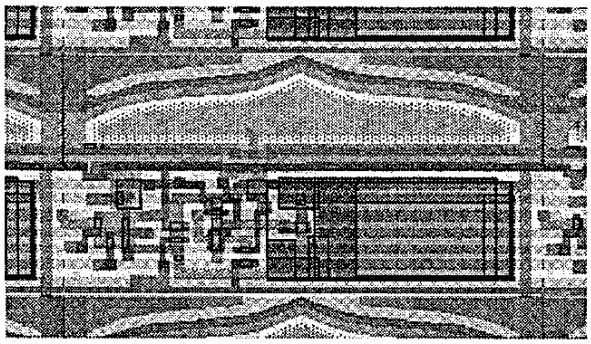

Figure 8: One pixel layout. The aspect ratio of each pixel is 2 x 1. The pixel measures 60 x 30μm in 0.5μ CMOS technology.

current that goes up and down as the laser travels across pixels. This differential current is brought to a one-per-row peak detector located on the side of each row.

Figure 7 shows the block diagram of the peak detector. A preamplifier turns the differential current into a differential voltage. This differential voltage exhibits zero crossings when the laser transitions from pixel to pixel. A first comparator detects these zero crossings and produces a digital waveform. A 90°-phase shifter produces a second differential voltage whose zero crossings correspond to the laser being positioned at the pixel centers. A second comparator produces a digital waveform that is in quadrature with the first digital waveform. The two digital waveforms are further combined to produce the latch signal for the one-per-row memory cell for the storage of the timestamp.

V. Chip Layout and Experimental Results

Figure 8 shows the layout of one pixel. The pixel aspect ratio is 2 x 1. After doubling resolution in the horizontal direction, the range map is sampled on a grid with a square aspect ratio (i.e., 1 x 1). The peculiar photodetector shape is intended to shape the photodetector signal as to produce steep zero crossings and pointy peaks as to aid in robust comparison at the comparators.

Figure 9 shows a range map captured by the sensor in 2ms, or 500 frames/second. Rather than using an explicit formula in Equation (1) for calculating the range based on the quantities captured by the sensor, we calibrated a lookup table to replace the Equation (1). The calibrated lookup table takes into account all geometrical variations that are present in the triangulation setup including: 1) variations of the base line with the mirror angle, and 2) non-constant angular velocity of the scanning mirror. For good laser returns we've measured speeds up to 2000 frames per second. This speed declines to about 60 frames per second for dark objects, which returned very little laser energy.

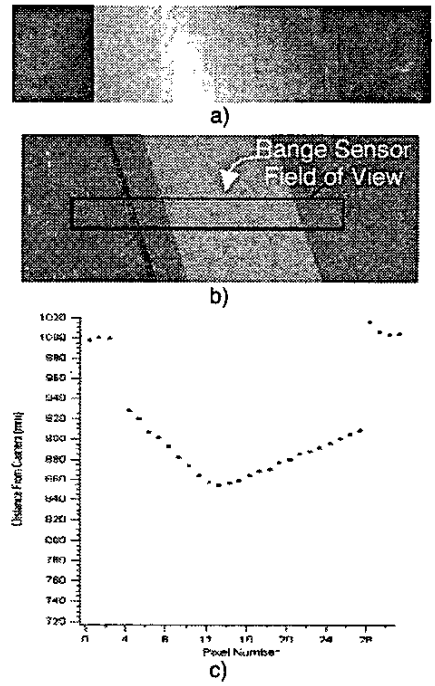

Figure 9: A captured range map of a cardboard box. The field of view for this range sensor prototype was 7 rows by 31 columns (16 photodetectors). Image a) is a gray coded range map captured by the range sensor. Image c) is the graph of range measurements for one row. Image b) is an appearance image captured by a CCD camera. The overlaid rectangle indicates the range sensor's field of view.

Acknowledgement

We would like to thank Mr. Terry Grayson for capturing the range map in Figure 9.

Refereces

[1] P.J. Besl, "Range Imaging Sensors," Res. Publ. GMR-6090, GM Research Labs., Warren, MI, Mar. 1988.

[2] Y. Sato, K. Araki, and S. Parthasarathy, "High speed rangefinder," in Optics, Illumination, and Image Sensing for Machine Vision II, SPIE, vol. 850, pp. 184-188, 1987.

[3] A. Gruss, L.R. Carley and T. Kanade, "Integrated Sensor and Range-Finding Analog Signal Processor," IEEE JSSC, vol. 26, No. 3, March 1991.

[4] A. Yokoyama, K. Sato, T. Yoshigahara, S. Inokuchi, "Realtime Range Imaging using Adjustment-Free Photo-VLSI – Silicon Range Finder," Proceedings IROS, pp. 1751-1758, 1994

[5] J. Lazzaro, S. Ryckebusch, M.A. Mahowald and C. Mead, "Winner-Take-All Networks of O(n) Complexity," in Advances in Neural Information Processing Systems Vol. 1, D. Tourestzky, ed., pp. 703-711, Morgan Kaufmann, San Mateo, CA, 1988.

[6] T. Delbrück, "Analog VLSI Phototransduction by Continuous-time, adaptive, logarithmic photoreceptor circuits," *CNS Memo No. 30*, California Institute of Technology, Pasadena, CA, 1994.

[7] S.-C. Liu, "Silicon retina with adaptive filtering properties," *Advances in Neural Information Processing Systems*, Vol. 10, Eds. Jordan, M.I., Kearns, M.J., and Solla, S.A., MIT Press, pp 712-718, 1998