A CMOS Expansion/Contraction Motion Sensor with a Retinal Processing Circuit for Z-motion Detection Applications

Chung-Yu Wu and Chien-Ta Huang
Department of Electronics Engineering and Institute of Electronics
National Chiao Tung University
1001 Ta-Hsueh Road, Hsinchu, 300, Taiwan
Email: peterwu@mail.nctu.edu.tw, chthuang.ee91g@nctu.edu.tw

***Abstract*— A design of CMOS expansion/contraction motion sensor based on the artificial retinal sensor array and digital correlation circuitry for real-time detecting object motion and velocity in the longitudinal +Z/-Z-directions (Z-motion) is proposed. To cope with the non-uniform nature in optic field of Z-motion, we compute the correlation between one set of output of retinal sensors and the delayed output of the other set of far-end sensors at progressive increasing distances across the chip for motion and velocity detection. Sweeping the correlation delay value in one signal path of the correlator, the chip can indicate the strongest correlation when the correlation delay is tuned to match the actual velocity. Behavior-level MATLAB simulations have been performed to successfully validate the functionity of the proposed motion detection method using expanding or contracting motion images that are incurred from an out-bound or in-bound moving object at a constant Z-velocity. The proposed motion sensor consists of 1448 sensor pixels placed on thirty-two evenly spaced divergent axes with 25 correlator outputs on each axis. Z-Motion and velocity detection is performed based on the maximum likelihood (ML) detection using the 25 correlators output of all 32 axes. The chip is designed in 0.35-um CMOS technology and the core occupies approximately 5600 x 5600 μm^2 with a fill factor of 11%.**

I. INTRODUCTION

Motion provides vital information to understand the environment. In many applications such as surveillance system, vehicle navigation, robot control systems, etc, motion sensors are used to detect and analyze the involved motion. One type of motion sensor, known as focal-plane motion sensor that integrates the function of image sensing and signal processing into a single chip, has drawn much intensive research effort recently. Focal-plane motion sensors have the advantages of fast processing time, small size, low cost, and low power dissipation. Most of the proposed focal-plane motion sensors [1]-[6] adopt the correlation-based algorithm to detect motion. Recently, a number of focal-plane sensors are reported to successfully utilize the pseudo-BJT-based (PBJT-based) retinal structure with Schmitt-trigger output for image perception. The combination of the PBJT retinal sensor and correlation-based circuitry has the benefits of robustness and compactness in CMOS VLSI-implementation [3].

Motion sensing and analysis generally can be performed based on the first-order motion information such as the optic flow. The optical flow of motion can be decomposed in three types Flow Components (EFCs), namely rotation, expansion (contraction), and shear motion [7]. In primate brains, these three EFCs are detected by three different medial superior temporal (MST) cells in the visual cortex. The MST neurons in the primates respond quickly to the EFCs motion stimulus. Related research [4] has been conducted to build a multi-chip system to discriminate wide-field spatial patterns of visual motion. However, the function of motion velocity calculation for general patterns is not implemented. Lately, focal-plane motion sensors that calculate the global velocities and directions of rotation [5] and shear [6] motion have been proposed. So far, motion sensors for expansion (contraction) detection have not yet been reported.

Inspired by the functionality of MST neurons in the primates for the detection of expansion (contraction) motion, a CMOS expansion/contraction motion sensor with a retinal processing circuit is firstly proposed. In this work, we design a CMOS expansion/contraction motion sensor that combines the artificial retinal sensor array and digital correlation circuitry. The functions of real-time calculation of global velocity and detection of expansion/contraction motion of general images are realized on a single chip. Mimicking the features of the expansion (contraction) selectivity of the MST neurons, the proposed motion sensor can selectively detect the expansion (contraction) motion by computing the correlation between the outputs of a group of sensors and the delayed outputs of another group on the same chip at chosen direction by a delay duration chosen according to the pre-selected Z-velocity. Furthermore, the correlation circuitry detects the Z-velocity by performing maximum likelihood detection based on the correlator outputs.

The rest of the paper is organized as follows: Section II describes the theory and mathematical model and the assumption of this work. Section III presents the method of detection method according to the model while Section IV is dedicated to circuit design and implementation. Section V shows the simulation results, followed by the final section, the conclusion.

II. THEORY AND ALGORITHM

The fundamental mathematical model for focal plane motion sensing, which is also common in computational vision and camera calibration theory, is the perspective projection equation.

1-4244-0921-7/07 $25.00

The perspective projection equation relates a point in 3-D space (X_i, Y_i, Z_i) point to a point 2-D sensory plane (x_i, y_i) that it projects onto (multiple to one mapping). Without loss of generality, we let x- and y- axes align with X and Y axes, respectively, Z-axis align with the focal axis, and the focal length being set to 1, perspective projection equation can be written as:

$$(x_i, y_i) = \left(\frac{X_i}{Z_i}, \frac{Y_i}{Z_i}\right) \tag{1}$$

To detect the motion in 3-D from the image on the planar sensor is difficult because the projection mapping results in that all pairs X(Y) and Z values that have the same X(Y) to Z ratio map to the same x(y). In other word, there is ambiguity between size and distance.

Since we are interested in the Z-velocity, we take the first time-derivative of the perspective projection equation. We get the Z-velocity and x- and y-velocity related by:

$$(\dot{x}_i, \dot{y}_i) = \left(\frac{\dot{X}_i Z_i - X_i \dot{Z}_i}{Z_i^2}, \frac{\dot{Y}_i Z_i - Y_i \dot{Z}_i}{Z_i^2}\right) = \left(\frac{\dot{X}_i - x_i \dot{Z}_i}{Z_i}, \frac{\dot{Y}_i - y_i \dot{Z}_i}{Z_i}\right) \tag{2}$$

The field pattern of first time-derivative of image x and y is in effect the optic field. This equation is a general formulation for describing image motion due to 3-D motion in which motion vectors in X-, Y-, and Z- direction are non-zero.

In this work, we assume all Z_i identical (= Z) without rotation and motion in X- and Y- direction, the time-derivative of projection equation can be reduced to:

$$(\dot{x}_i, \dot{y}_i) = \left(\frac{-x_i V_z}{Z}, \frac{-y_i V_z}{Z}\right) \tag{3}$$

In computational vision literatures, a theoretical approach for solving V_z is to track the motion/velocity of a number of sample points (indexed by *i*) on the sensor plane and solve the following minimization problem of the weighed sum of square error using location and velocity of x and y measured for the sampled points. The minimization problem is formulated as follows:

$$\arg\min\left(\sum_i [w_{xi}(\dot{x}_i Z + x_i V_z)^2 + w_{yi}(\dot{y}_i Z + y_i V_z)^2]\right)$$

where $\mathrm{w_{xi}}$'s $\mathrm{w_{yi}}$'s are weighting factors,

$x_i, \dot{x}_i, y_i, \dot{y}_i$ are known measured quantities

$\mathrm{V_z}$ and Z (= $\int \mathrm{V_z} dt$) are unknown to solve

However, the computation requirements for this algorithm are too much to fit into a compact single analog integrated circuit that in general has very little computation power, much less that measuring instantaneous velocity in x and y is typically regarded a noisy process.

In this work, we assume the moving object is a thin rigid-body object taking a single Z-coordinate and moving in purely Z-direction at constant velocity. We also consider that Z-velocity is the global velocity to be detected because the optic flow in the sensor plane primarily results from the motion of object. To resolve the distance-size ambiguity, we presume to detect when the object arrives at a reference distance (Z_0) away from the sensor by some calibration means beforehand or after. By this way, we effectively impose a constraint that relates Z_0 and V_z with time ΔT:

$$Z = Z_0 + V_z * \Delta T \tag{4}$$

Unlike the optic field of planar motion in translation or rotation, which is uniform in either x-/y- direction or angular direction, the optic field of the Z-motion is non-uniform and time-varying. Because of non-uniform and time-varying nature of the image on the sensor plane, we prefer detecting the average velocity (that is, compute $V_z = \Delta Z/\Delta T$) to the instantaneous velocity of the image motion. The Z-velocity detection by computing average velocity based a correlation-based method is described in Section III.

III. MOTION DETECION METHOD

The rationale of the design of the correlation-based method comes directly from the perspective projection equation. When an object is moving from a location of z unit distance to 1 unit distance from the sensor along the Z-axis, an image point at radial distance or polar coordinate *r* ($r = \sqrt{x^2 + y^2}$) away from the origin will move to a point at *z* times of *r* away from the origin at the same angle θ. Fig 1 illustrates the case for *z*=2.

Let's denote the output of the sensor at polar coordinate (r, θ) by $p_\theta(r, t)$ as a function of time t. The subscript θ distinguishes the sensors located in different angular directions. A reference distance Z_0 is chosen at our choice. Suppose that in duration *T,* the object travels from distance of $z*Z_0$ to Z_0 (z > 1 for expansion case while 0 < *z* < 1 for contraction motion). Then, if an image point currently (at time *t*) is located at *r*, it was at $r*z^{-1}$ at (*t-T*) before, where T=($|1-z^{-1}|*Z_0/V_z$).

Ideally, we can write a correlation function by integrating the correlation between one pixel $p_\theta(r, t)$ and the delayed output of pairing pixel $p_\theta(r*z^{-1}, t-Dt)$ with a correlation delay *Dt* along the polar coordinate *r* at each angle θ as:

$$c_\theta(t) = \frac{1}{\int dr} \int equ(p_\theta(r,t), p_\theta(r \cdot z^{-1}, t - Dt))\, dr \tag{5}$$

where correlation between two sensor pixels is computed by *equ* (*in1*,*in2*) that is the indicator function of two-input equivalence. When both inputs of the *equ* function have the same polarity (both are logic high or low), the function gives 1 (a logical high signal), otherwise, yields 0 (logical low). The output range of the correlation function is between 0 and 1, where 1 indicates "perfect correlation" when correlation delay *Dt* is exactly tuned to ($|1-z^{-1}|*Z_0/V_z$).

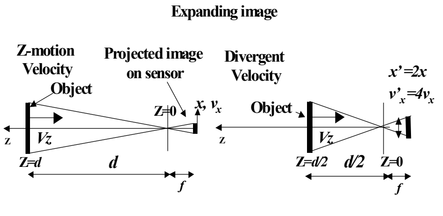


Figure 1. The relationship between (expansion/contraction) x-/y- velocity *v* and Z-motion velocity *Vz,* where $\Delta d = d/2$ is the travel distance of an object and *d* the distance between an object and the motion sensor.

It should be easy to show that if the correlation delay Dt is chosen to be $(|1\text{-}z^{-1}|*Z_0/V_z)$, the output of sensor located at r should be in correlation with $(r*z^{-1})$ for all r. And let correlation function output $C_\theta = c_\theta$ (t) be evaluated when Z=Z_0.

In the realistic case, sensor pixels have finite area, and it is desirable that sensors are placed on the multiples of pitches with integer index (r = k*$pitch$). Therefore, we rewrite the correlation function in discrete domain by replacing r with k.

$$c_\theta(t) = \frac{1}{\sum_k 1}\sum_k equ(p_\theta(k,t), p_\theta(z\cdot k, t-Dt)) \qquad (6)$$

Care must be taken in evaluating the correlation function in discrete domain. We only sum up pair of sensor pixels whose index k and $z*k$ are integers by careful design. We also require that during the detection interval, the object's image be entirely covered by the sensor plane. Moreover, we exclude trivial/monotone image motion patterns such as a sequence of all-bright or all-dark images because these patterns always yield perfect correlation.

The notion of correlation pair consisting of two sensor and two correlators for expansion and contraction correlation is depicted in Fig. 2. "CF" and "PF" stand for the current field and previous field registers. At the system level, the outputs of all correlators are aggregated all over the sensor plane to determine motion in expansion or contraction and the Z-velocity in ensemble (combining correlation outputs at all angle). By adjusting the correlation delay Dt set for the delayed input of correlators, we can make the sensor chip sensitive to different Z-velocity. As a result, when the object is moving at some constant Z-velocity, if we plot the averaged correlator function output ($\frac{1}{M}\sum_0^M C_{\theta k}$ over all angles) against the delay value of the correlator, we can plot a curve that attain the peak value at the delay corresponding to desired Z-velocity. The maximum likelihood detection for the motion and velocity detection is to determine if the peak of correlation output achieves a certain threshold value and the curve is sharp enough to distinguish the peak point from others. Then we can decide that the object is moving at a constant velocity (determined by Z_o and $Dt = (|1\text{-}z^{-1}|*Z_0/V_Z)$) in the +Z or –Z direction.

IV. CHIP DESIGN AND IMPLEMENTATION

The image acquisition and preprocessing in a motion sensor can be performed by a retinal processing circuit. The retinal processing circuit emulates reception parts of biological functions of the retinal cells.As the real retina, the retinal processing circuit has similar advantages of high dynamic range, edge enhancement, and noise immunity. In the proposed structures [2]-[3], [5], [8], the pseudo-BJT-based (PBJT-based) retinal structure is rather compact and fully compatible with the CMOS technology. The photoreceptor senses the input images while the retinal smoothing network extracts the local spatial and temporal averages of the incident images. The output currents of the retinal smoothing network and the photoreceptor are sent to the adaptive current Schmitt trigger, which amplifies the difference between the two input currents with hysteresis. The output of the adaptive current Schmitt trigger is a voltage signal which is further amplified to VDD or GND by an inverter. The output image of pixels becomes a black-and-white image. In the black region, the output voltage of the retinal processing circuit will be GND, which is logic zero, whereas in the white region, the output voltage will be VDD, which is logic one. In the real chip design, we intend to detect z-motion for an object travels from $z*Zo$ to Zo with z=2 and ½ for expansion and contraction motion, respectively. By this choice, the correlation pairs will the same for both motions.

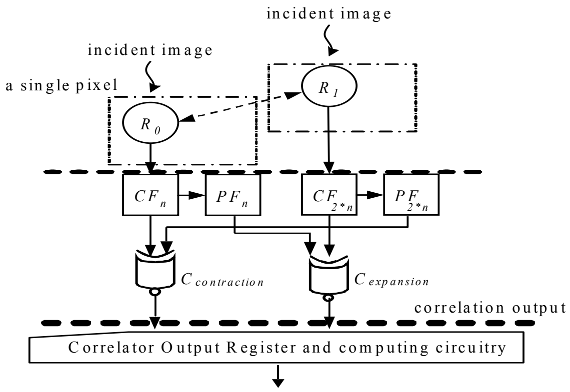


Figure 2. Conceptual structure of the adapted correlation-based algorithm for expansion/contraction motion detection.

Fig. 3 shows the architecture of the proposed expansion/contraction motion sensor, which includes the sensing array, a group of digital computation circuits, and a group of digital scanning circuits. If we want to maintain error than 5% when one out of all correlator is in an error, the number of the correlators has to be at least 20. We choose the number of correlation pair to be 25 for over-design and thus number of pixels in an axis is 50. Because diameter of each retinal cell is about 46 µm, the pitch of each axis is chosen to be 50 µm in order to have high fill-factor. However, for that reason, some cells in the inner ring have to be dropped due to placement and routing constraints. In the real implementation, only 4 out of 32 axis have 49 cells in an array, 12 axes have 47 cells, and the remaining 16 axes have only 43 cells. The real chip implementation of sensor array consists of 1488 sensor pixels.

At the chip level, the outputs of the correlators are read from a chain of scan registers controlled by a control signal. The chip is initially reset and started by a control signal. At the end, the correlation is calculated and collected by a host computer. Later, we analyze the curve of correlation function output generated by repeatedly sweeping correlation delay to perform Maximum Likelihood Detection for Z-velocity. Fig. 4 is the schematic of the PBJT-based retinal processing circuit. It consists of an isolated PNP pseudo-BJT used as photoreceptor, a smoothing NPN pseudo-BJT with adjustable N-channel MOSFET resistors used to form the retinal smoothing network, an adaptive current Schmitt trigger, and an inverter. The retina cell is modified from [8] with grounding P-substrate in the photodiode (PD) for improving current efficiency and reducing device dimensions. By our previous test chip, the size of PD can generate photo current of 0.2nA under 95 lux and 20nA under 2010 lux

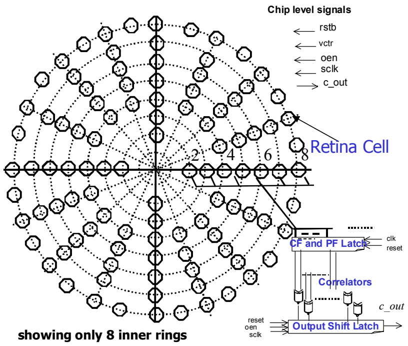
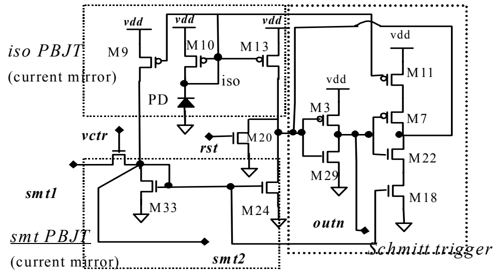


Figure 3. Architecture of the proposed expansion/contraction motion sensor



rst *is inverted rstb input* ***outn*** *is prior output(inverted)*

Figure 4. The PBJT-based retinal processing circuit modified from [8]

V. SIMULATION RESULTS

In behavior simulations in MATLAB, we create sequences of expanding/contracting images in resolution 1023 by 1023 pixel for testing. Then, we emulate the function of retinal processing circuitry by sampling each image in sequence in 50*32 small windows; each has 9 (3 by 3) pixels (effective sensing area). The ratio of window dimension (=3) to the whole sensor plane (=1023) is chosen to approximate the ratio of the real sensor pixel in the real sensor core area. The content of each pixel in 25-pixel window is summed and compared to a threshold (set to half of the full-scale). If larger than the threshold, the sensor outputs 1; otherwise outputs 0. In effect, the array of sensor window outputs a 'black-and-white' image.

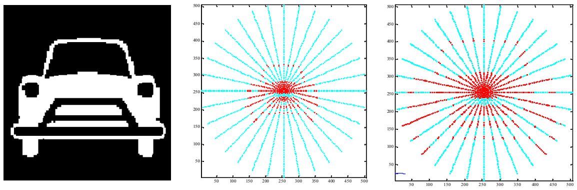
Fig. 5 shows input image for simulation and snapshot of sensor pixel output when object moves at $2Z_0$ and Z_0. Fig. 6 illustrates the behavioral simulation result (Maximum likelihood detection) by plotting the curve of the averaged correlator function output (averaged over 32 axes) against the various delays set between the correlation pair. By this, we validate the function of the proposed correlation-based Z-motion using assort of image sequences by choosing different sample rates. In transistor-level SPICE simulations, we provide the stimulus in currents in place of photodiodes and get same circuit characteristics as reported in [8].

VI. CONCLUSION

A CMOS expansion/contraction focal-plane motion sensor that combines a retinal processing circuit and the correlation-based detection method is implemented for real-time Z-motion detection applications. By imitating the expansion (contraction) selectivity of the MST neurons, the proposed motion sensor is designed to selectively detect the expansion (contraction) motion. The proposed circuit is the first in the motion-sensor research area that can detect moving images in expansion or contraction motion due to general rigid-body object moving at a constant speed in the Z-direction. The digital correlation circuit in the focal-plane also distinguishes itself from other correlation circuitry for it is designed to cope with non-uniform and time-varying optic field pattern. In the future, more specified functionality could be adaptively designed on the chip according to desired motion detection applications such as moving robot and automotive systems.

VII. ACKNOWLEDGMENT

The authors would like to acknowledge fabrication support provided by National Chip Implementation Center(CIC).



input pattern

Sensor pixel outputs (red dots) by sampling motion image when object moves at distance 2*Zo and Zo

Figure 5. input test image and snapshot of simpled sensor pixel output in Matlab simulation

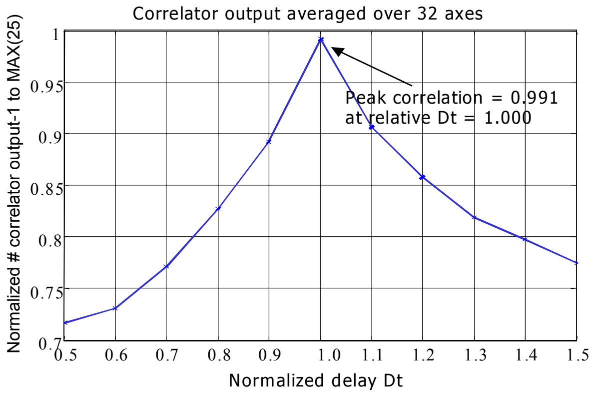


Figure 6. The averaged correlation function output (over 32 divergent directions) versus various delay (normalized to desired delay).

REFERENCES

[1] C. M. Higgins, R.A. Deutschmann and C. Koch, "Pulse-based 2-D motion sensors," IEEE Transactions on Circuits and Systems II, vol. 46, no. 6, pp 677-687, June 1999.

[2] Hsin-Chin Jiang and Chung-Yu Wu, "A 2-D velocity- and direction-selective sensor with BJT-based silicon retina and temporal zero-crossing detector," IEEE Journal of Solid-State Circuits, vol. 34, no. 2, pp. 241-247, Feb. 1999.

[3] Chung-Yu Wu and Kuan-Hsun Huang, "A CMOS focal-plane motion sensor with BJT-based retinal smoothing network and modified correlation-based algorithm," IEEE Sensors Journal, vol. 2, no. 6, pp. 549-558, Dec. 2002.

[4] C. M. Higgins and S. A. Shams, "A Biologically Inspired Modular VLSI System for Visual Measurement of Self-motion," IEEE Sensors Journal, vol. 2, no. 6, pp. 508-528, Dec. 2002.

[5] Kuan-Hsun Huang, Li-Ju Lin, and Chung-Yu Wu, "Analysis and Design of a CMOS Angular Velocity- and Direction-Selective Rotation Sensor with a Retinal Processing Circuit," IEEE Sensors Journal, vol. 4, no. 6, pp. 845-856, Dec. 2004.

[6] Chung-Yu Wu, Wen-Chin Hsieh, and Cheng-Ta Chiang, "A CMOS Focal-Plane Retinal sensor Designed for shear Motion Detection," in Proc. of IEEE International Conference on Electronics, Circuits and Systems, ICECS'04, pp. 141-144, Dec. 2004.

[7] G. A. Orban and H. –H Nagel (eds.), Artificial and Biological Vision Systems, Berlin: Springer-Verlag, 1992.

[8] Chung-Yu Wu and Cheng-Ta Chiang, "A Low-photocurrent CMOS Retinal Focal-plane Sensor with a Pseudo-BJT Smoothing Network and an Adaptive Current Schmitt Trigger for Scanner Applications," IEEE Sensors Journal, vol. 4, no. 4, pp. 510-518, Aug. 2004