A SPACE-VARIANT SENSOR STRUCTURE FOR COMPLEX TARGET DETECTION AND TRACKING

David Claveau and Chunyan Wang

Department of Electrical and Computer Engineering
Concordia University, Montreal, Que., Canada H3G 1M8
{cd_clave,chunyan}@ece.concordia.ca

ABSTRACT

Visually tracking a real-world target with an image sensor is widely applicable in robotics and human-computer interaction but is a computationally intensive task. Smart image sensors can make use of integrated motion detection circuits for target tracking, but algorithms for tracking complex, real-world targets are usually not suited to smart sensor implementations. We propose a sensor structure for tracking targets that have complex spatial frequency patterns and varying velocities. In this structure only basic motion detection circuits are used, however the distance between sensing elements is varied in order to vary the filtering properties of the system to be sensitive to a broad range of spatial frequencies and velocities. Simulation results show that it is possible to be selectively sensitive to both the different spatial frequencies of a target and to its varying velocity.

1. INTRODUCTION

Visually tracking a real-world target with an image sensor is an important part of many applications in fields such as human-computer interaction [1] and mobile robotics [2]. In such systems, the image sensor is part of a visual-feedback control loop as shown in Fig. 1. Information about the target must be extracted from the images acquired by the sensor. This information is used by the control system to keep the sensor aligned with the target. Even at modest resolutions a large amount of data must be processed and transferred through the system. To address this problem researchers have begun to use smart image sensors that have motion detection circuits integrated directly on the array of photoreceptors. This allows for parallel processing of the image data and the generation of a control signal on the same chip, as indicated by the shaded area in Fig. 1. But the difficulty of tracking real-world targets that have complex patterns and move with varying velocity remains. The solution generally requires a computationally complex algorithm that is not suited to a smart sensor architecture.

There has been some research reported on smart image sensors that use motion detectors for target tracking. A 1-D sensor array with integrated motion detection circuits was described in [3]. The sensor was used in a complete sensory-motor control system, however the issues of tracking complex targets were not explored. A related system reported in [4] estimates self-motion using four different arrays of motion detection circuits which can be tuned to different orientations, speeds or other properties. This offers the possibility to deal with complex targets. The complete system requires multiple chips for the different arrays and some nontrivial coordination between them.

Typically, sensor arrays have a uniform sampling grid resulting in a uniform resolution across the sensor. The term space-variant [5] refers to a variation in the resolution of an image sensor across its surface, as in the human vision system. Much of the published research on space-variant or 'foveated' sensors has focused on their application to data reduction for efficient image transmission and as a way to achieve a wide field-of-view along with a region of high resolution. An example of a foveated sensor for tracking is described in [6]. It has a central region of uniformly high resolution and a peripheral region of lower resolution. Motion detection is only performed in the central region

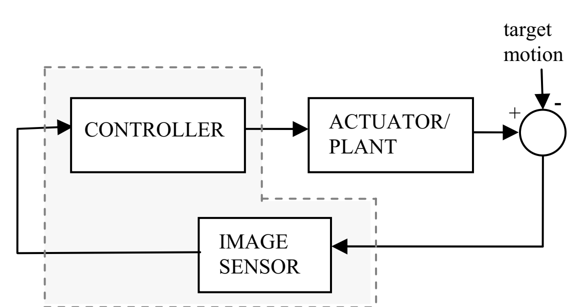

Figure 1. Visual-feedback control loop with image sensor. Shaded area indicates how a smart sensor can potentially perform both sensing and control.

across uniformly spaced photoreceptors, while spatio-temporal edge detection is performed in the periphery for target acquisition.

In this paper we propose a space-variant sensor structure for tracking complex targets using only basic motion detection circuits. In this structure, the distance between sensing elements is varied in order to vary the filtering properties of the motion detectors to be sensitive to both the different spatial frequencies of a target and to its varying velocity.

2. THE PROPOSED SYSTEM

The goal of a visual target tracking system is to keep the optical axis of the image sensor aligned with a moving target. For smart image sensors, motion information is often extracted using some form of correlation-based motion detection. A simple and commonly used circuit is the Reichardt elementary motion detector (EMD) [7]. A block diagram of the basic EMD is shown in Fig. 2. The photoreceptors are separated by a distance, d. If a target with a sinusoidal pattern of amplitude 1 and spatial frequency f_x passes over them with velocity, v, then the average response over time can be expressed as a function of spatial and temporal frequency:

$$\overline{R}(f_x, f_t) = \frac{1}{2\pi\tau} \cdot \left[\frac{f_t}{f_t^2 + 1/(2\pi\tau)^2} \right] \cdot \left[\sin(2\pi \cdot f_x \cdot d) \right] \quad (1)$$

in which $f_t = f_x v$, the temporal frequency induced by the motion of the target. From the first term we see that the response will peak at a temporal frequency of:

$$f_{t,peak} = \frac{1}{2\pi\tau}. \quad (2)$$

The sinusoidal term tells us that the response will peak for a spatial frequency of:

$$f_{x,peak} = \frac{1}{4d}. \quad (3)$$

The value of $f_{x,peak}$ depends on the spacing, d. In order to see the shape of individual response curves for targets with particular spatial frequencies, a plot of four different spatial frequencies is shown in Fig. 3. For this plot $\tau = 0.1$ sec. and $d = 1$, which is the smallest d we consider, and we make this a basic unit that we can use to express the velocity. For example, the v_{peak} is 6.4 units/sec in Fig. 3. For a fixed d and τ, our ability to detect a moving target and estimate its velocity is limited to a narrow range. The sensitivity of the response to a change in velocity is very high around $f_{x,peak}$ and v_{peak} but decreases rapidly for the spatial frequencies around it. The locations of $f_{x,peak}$ and v_{peak} depend on the choice of d, as seen in equations (2) and (3).

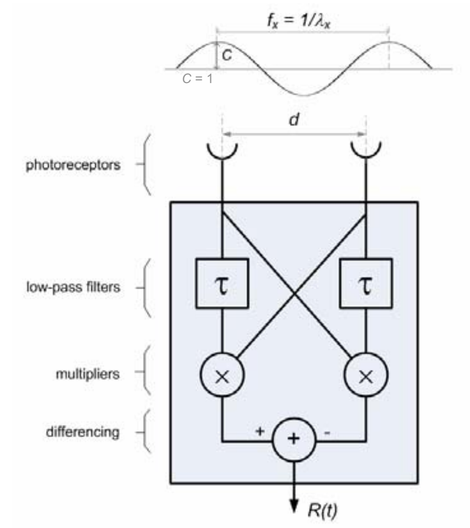

Figure 2. Block diagram of a simplified Reichardt EMD.

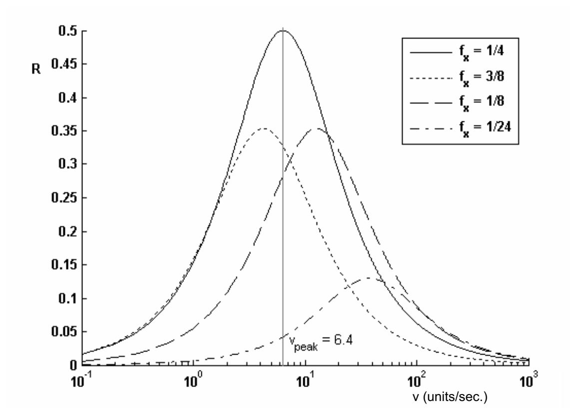

Figure 3. EMD response to drifting sinusoidal gratings ($d = 1$, $\tau = 0.1$sec.) with curves for different spatial frequencies superimposed.

Now consider how an EMD response curve is shifted when d is changed. Fig. 4 shows a contour diagram of two EMDs ($d = 1$, $d = 2.7$) and the frequency and velocity ranges they cover for a response above a certain threshold. In this particular case the threshold is $R_{TH} = 0.35$. The frequency and velocity ranges are shown with dashed lines. As d is increased, v_{peak} increases while $f_{x,peak}$ decreases. In this case the velocity ranges overlap but a response from a particular EMD will still indicate motion of a target within a specific velocity and spatial frequency range. The velocity and spatial frequency bandwidths can be shifted to different ranges depending on the value of d.

An array can be designed to have EMDs with different photoreceptor spacings, similar in principle to the system discussed in [8]. Consider a one-dimensional sensor array that can move left/right to track a target that enters its field-of-view from the left. Assuming a uniform array, as shown

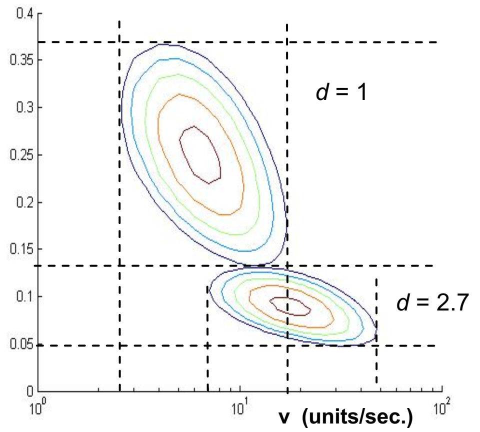

Figure 4. Contour diagram of the response to a drifting sinusoidal grating for two different EMDs with a threshold of $R_{TH} = 0.35$.

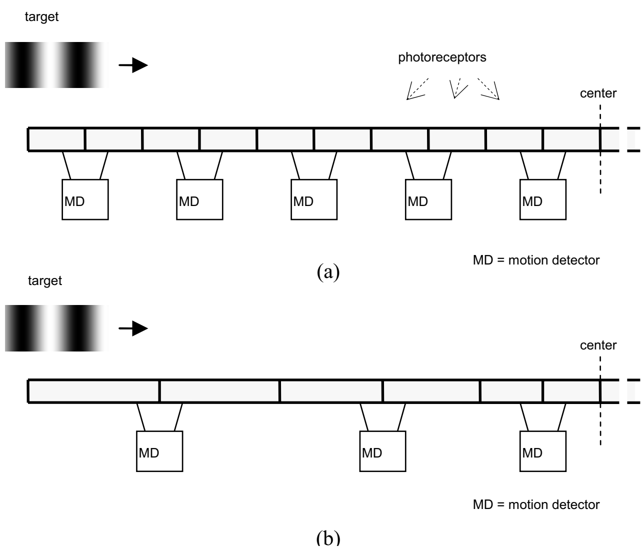

Figure 5. Uniform array (a) and a space-variant array (b) with a target moving into the field-of-view from the left.

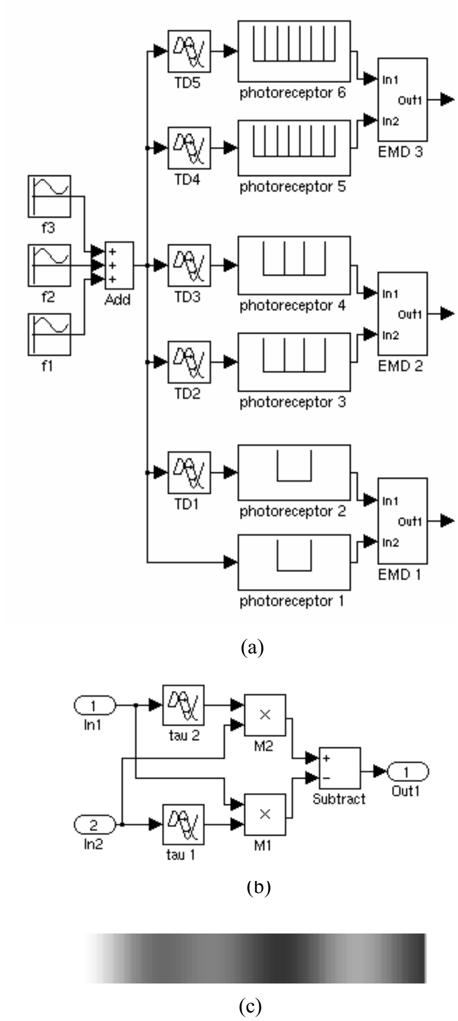

Figure 6. Simulink models for the space-variant array (a) and the individual EMDs (b), along with a visual representation of the simulated target in (c).

in Fig. 5 (a), any motion exhibited by the target, significant or not, will be detected and the sensor will attempt to track this motion. In the space-variant array shown in Fig. 5 (b), the wide photoreceptors of the periphery will only detect the lowest spatial frequencies of the target. The higher spatial frequencies will only be detected as the target becomes more aligned with the center of the sensor. This enables the sensor to initially filter out motion that might cause it to lose track of the target because of a small but sudden change in velocity. There are two main advantages to this approach:

1) By varying *d* across the array, a sampling structure can be defined along with a threshold in such a way as to enable detection over a larger range of velocities and spatial frequencies than for a uniform array.
2) A sampling structure can be defined to separately detect motion indicated by the low and high spatial frequencies of a target. A coarse-to-fine tracking strategy can then be used, effectively increasing the precision of the system as the target becomes more aligned with the sensor.

3. SIMULATION AND RESULTS

To investigate the proposed structure, a simulation of a space-variant array was developed in Matlab/Simulink. To simulate the motion of a target that has a complex spatial frequency pattern, the block diagram in Fig. 6 (a) was used. The idea is to use a bank of moving average filters to represent the photoreceptors of different sizes. Then the moving target can be simulated with a temporal sinusoidal signal composed of three frequencies in a 1:2:4 ratio. The photoreceptor sizes are also in a 1:2:4 ratio and are paired up to feed into EMDs which are simulated with the block diagram shown in Fig. 6 (b). The sinusoidal grating in Fig. 6

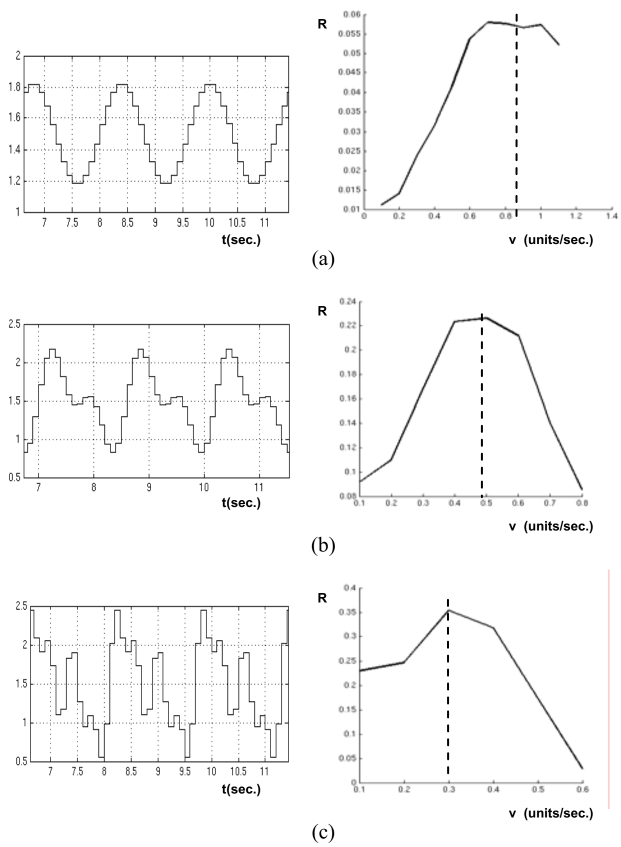

Figure 7. Photoreceptor output (left) and EMD response (right) for EMD 3 (a) EMD 2 (b) and EMD 1 (c) corresponding with Figure 6.

(c) shows how the target would appear visually. The sizing of the photoreceptors essentially determines the '*d'* for each EMD. The ratio of the photoreceptor sizes corresponds with the component frequencies of the target. This reflects some *a priori* knowledge of the spectrum of the target's spatial frequencies. Depending on the application, a sensor array can be designed to be selectively sensitive to these frequencies. As shown in Fig. 7 on the left, the outputs of the photoreceptors have separated the spatial frequencies of the target. The largest photoreceptor outputs only the changes caused by the motion of the lowest spatial frequency in the target, as shown in Fig. 7 (a). In Fig. 7(b) and (c) the smaller photoreceptors output changes due to motion of the higher spatial frequency components. By moving the target across the array at different velocities we can plot the average response of the EMDs as shown on the right of Fig. 7. The average response of the EMD with the largest photoreceptors is shown in Fig. 7 (a) while the next smaller ones are shown in (b) and (c). Each EMD has a particular range of velocities for which it is more sensitive. The EMD in (a) is sensitive to higher velocities than (b) and (c). In each case it is possible to threshold these responses to more narrowly define the velocity ranges. Such an array can be designed to be sensitive to a large velocity range and to be able to distinguish between different velocities relative to the sensor. Thus, the distribution of EMDs can be tailored to detect the various velocities and spatial frequency components of a complex, real-world target. Tracking is expected to improve without incurring the cost of extra hardware or more complicated circuitry.

4. CONCLUSION

In this paper a space-variant sensor structure has been proposed. Only basic motion detection circuits are used, however the distance between sensing elements is varied in order to vary the filtering properties of the system to be sensitive to a broad range of spatial frequencies and velocities. Depending on the distribution and sizes of the photoreceptors, the EMDs can be separately tuned to the spatial frequency components of the target. They are also sensitive to the different velocity ranges of the target. This filtering can be tailored to different applications as long some knowledge of the spectrum of the target exists. The filtering can then be used to improve detection and tracking by extracting the most useful information about the target's motion depending on its position and velocity relative to the sensor.

11. REFERENCES

[1] Wren, C.R., Azarbayejani, A., Darrell, T., and Pentland, A.P., "Pfinder: Real-Time Tracking of the Human Body," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, No. 7, July 1997.

[2] Feyrer, S. and Zell, A., "Tracking and Pursuing Persons with a Mobile Robot," International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, p. 83, 1999.

[3] R.R. Harrison and C. Koch, "A silicon implementation of the fly's optomotor control system," Neural Computation, 12: 2291-2304, 2000.

[4] C.M. Higgins, and S.A. Shams, "A Biologically-Inspired Modular VLSI System for Visual Measurement of Self-Motion," IEEE Sensors Journal special issue on Integrated Multi-Sensor Systems and Signal Processing, pp. 508--528, December 2002.

[5] Schwartz, E.L., Greve, D.N., Bonmassar, G., "Space-Variant Active Vision: Definition, Overview and Examples," *Neural Networks(8)*, No. 7-8, 1995, pp. 1297-1308.

[6] R. Etienne-Cummings, J. Van der Spiegel, P. Mueller and M.Z. Zhang, "A Foveated Silicon Retina for Two-Dimensional Tracking", IEEE Trans. Circuits and Systems II, Vol. 47, pp. 504-517, June 2000.

[7] W. Reichardt, in Sensory Communication, ed. W.A. Rosenblith, pp. 303-317, MIT Press, Cambridge, Massachusetts, 1961.

[8] Claveau, D. and Wang, C., "A Spatial Variance Approach to Target Tracking with Sensor Arrays," IEEE International Symposium on Circuits and Systems, 2005.