# Space-Variant Motion Detection for Active Visual Target Tracking

David Claveau and Chunyan Wang
Electrical and Computer Engineering
Concordia University
Montreal, Canada

**Corresponding Author:**

Chunyan Wang
chunyan@ece.concordia.ca
(514) 848-2424 ext. 3120 (voice)
(514) 848-2802 (fax)

Dept. of Electrical and Computer Engineering
Concordia University
S-EV005.139, 1515 St. Catherine, West,
Montreal, Quebec CANADA  H4G 2W1

## Abstract

A biologically inspired approach to active visual target tracking is presented. The approach makes use of three strategies found in biological systems: space-variant sensing, a spatio-temporal frequency based model of motion detection and the alignment of sensory-motor maps. Space-variant imaging is used to create a 1-D array of elementary motion detectors (EMDs) that are tuned in such a way as to make it possible to detect motion over a wide range of velocities while still being able to detect motion precisely. The array is incorporated into an active visual tracking system. A method of analysis and design for such a tracking system is proposed. It makes use of a sensory-motor map which consists of a phase-plane plot of the continuous-time dynamics of the tracking system overlaid onto a map of the detection capabilities of the array of EMDs. This sensory-motor map is used to design a simple 1-D tracking system and several simulations show how the method can be used to control tracking performance using such metrics as overshoot and settling time. A complete 1-D active vision system is implemented and a set of simple target tracking experiments are performed to demonstrate the effectiveness of the approach.

*Keywords*: active vision, sensory-motor coordination, space-variant imaging, visual motion tracking

## 1. Introduction

A visually guided task such as target tracking presents many challenges to the designer of a robot vision system. These challenges appear to have been met by nature, as in the case of humans. We can easily track a moving target with our eyes but have difficulty designing this ability into our robots. In part it is difficult because an effective target tracking system should be able to detect a large range of target speeds while still being able to precisely detect the fine movements of a target. This combination of large range of operation and high precision is difficult to achieve. Another source of difficulty is the need for sensory-motor coordination. This means that data must be extracted from an image signal in order to produce a useful action in a real-time control loop. This is challenging because it is often difficult for the designer to visualize how the sensing capabilities of a system relate to the motor capabilities. One way of addressing these challenges is by taking inspiration from nature. In this paper we consider the design of a bio-inspired active visual target tracking system that uses a visual-feedback control loop to keep an image-sensor/camera aligned with a moving target. Our approach makes use of three strategies found in biological vision systems: i) space-variant or foveal sensing, ii) spatio-temporal frequency based models of motion detection and iii) aligned sensory-motor maps.

Biological vision systems of many animals make use of a distinctive arrangement of photoreceptors which consists of a region of high resolution, the fovea, surrounded by a region of progressively lower resolution. Inspired by this, computer vision uses the term space-variant [1] to refer to a variation in the resolution of an imaging system across its field-of-view.  With a foveal arrangement, an active vision system can direct the fovea at a point of interest for precise imaging while maintaining a wide field-of-view. Much of the published research on space-variant or 'foveal' imaging has focused on this feature [2][3]. There are examples in the literature of active visual tracking systems that make use of space-variant sensing but almost all of these systems focus on the reduction of data that must be processed or on the computational advantages of a log-polar mapping to process images more efficiently [4][5][6].  However, a foveal pattern of processing is also present after the retina, along the visual pathway, with the variation of receptive field sizes of subsequent neurons [7]. This suggests that space-variance plays an important role in a variety of visual processing such as motion detection.

Visual motion detection has been well studied and several different models and algorithms exist [8] such as the gradient-based methods [9][10] and the spatio-temporal frequency based methods which include the motion energy models [11] and the correlation models [12]. Experimental studies have shown that biological vision systems likely use spatio-temporal frequency based techniques [13]. These usually consist of an array of tuned spatio-temporal frequency filters that are selectively activated by different types of visual motion. This can be an advantage for higher-level tasks like target tracking which can use this selective activation to better form a motor control signal. A simple and familiar example of a spatio-temporal frequency based model is the Reichardt elementary motion detector (EMD) model [12]. The EMD detects motion locally between two points and makes use of only a small number of simple operations. Versions of it have been used in smart sensor implementations [14][15]. One drawback is that an array of

uniform EMDs will be limited to detection over a very specific range of target speeds and spatial frequency patterns. One way to overcome this is to create an array or several arrays of differently tuned EMDs. Higgins and Shams [16] have proposed a system that uses four arrays of EMDs distinguished by differences in EMD orientation, speed tuning, spatial frequency tuning or other properties. Such an arrangement is powerful but also requires significant computational resources by using four complete and separate arrays. What is needed is an effective way to incorporate the differently tuned EMDs into a single array.

Once motion information has been extracted from an image signal, the task that remains is to produce an appropriate control signal to drive the vision system in such a way as to track the target. This sensory-motor coordination is difficult to design and is often designed in a task specific way. For example, systems by Ruffier and Franceschini [17] and Harrison and Koch [14] describe visual control loops that use the EMD to measure motion for flight control applications. What is needed is a tool to help visualize how the sensing and motor functions interact. Such a tool could be based on research in neurophysiology which has shown that sensory-motor regions of the vertebrate brain, such as the superior colliculus, receive inputs from the eye in a retinotopic map that is aligned with a 'motor map' of neurons that project to brainstem areas which then control eye and head movement [18].

Using these these bio-inspired strategies, this paper describes how a simple active visual target tracking system can be designed to track a target over a wide range of speeds while still being able to precisely track its fine movements. Section II describes in detail how space-variant sensing can be used to create an array of EMDs that are tuned to provide both range and precision for target tracking. Section III describes how an effective control signal can be computed by aligning a plot of the motion detection capabilities with a plot of the tracking dynamics. In Section IV an experimental system is described along with the results of target tracking experiments.

## 2. Space-variant motion detection

In this section we investigate how the idea of space-variant imaging can be used to create an array of tuned EMDs. A simple Reichardt EMD model is used because it is simple, commonly used and makes it easy to see the parameters that are adjustable and their effects on the response. In fact the results should be applicable to any tunable spatio-temporal frequency based model of motion detection.

### 2.1. Tuning a Reichardt EMD

A simple Reichardt EMD model is shown in Fig. 1. It has two input signals, $p_1$ and $p_2$, acquired from two points separated by a distance $\Delta x$. The inputs are assumed to come from photodetectors and represent the image irradiance at two adjacent points in the image plane. The input $p_1$ is delayed by $\Delta t$ and is multiplied with the undelayed input $p_2$, producing a response that is sensitive to rightward motion. The same delay-and-multiply operation is performed for the opposite direction. The outputs of the two sides are subtracted to reduce the response to static patterns and full-field flicker. For the configuration in Fig. 1 the output will

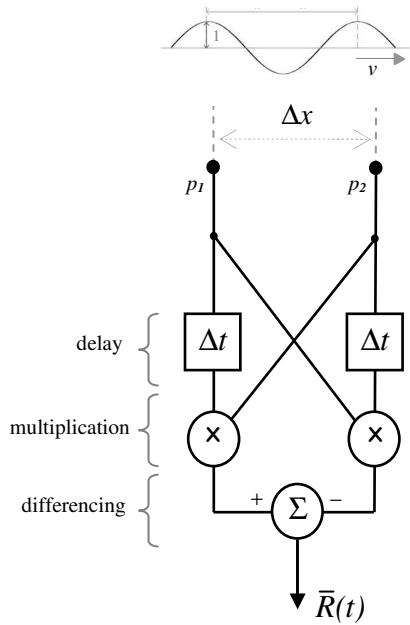be positive for rightward motion and negative for leftward motion.



Fig. 1.  Block diagram of a simplified Reichardt EMD showing the two
adjustable parameters: $\Delta x$, the distance between inputs and $\Delta t$ the delay.
The input sinusoid, at top, could represent a target passing over the
EMD.

   The two parameters that can be adjusted in the simplified EMD are the distance, $\Delta x$, between the two
points where the inputs are acquired, and the delay, $\Delta t$. To see how they affect the response of the EMD to
motion it helps to assume that the input to the EMD is a zero-mean sinusoid that has unit amplitude and a
spatial frequency of $f_x = 1/\lambda_x$ as shown at the top of Fig. 1. This assumes that a phototransduction stage is
able to produce such a signal from the image of an actual target and allows us to ignore for now the effect
of contrast on the response. If the sinusoid is shifted to the right with velocity $v$ then a temporal frequency
of $f_t = f_x v$ will be induced at each input. The shifting sinusoid could represent a target passing over the
EMD. If the delay is modeled as the inherent phase delay of a first-order low pass filter with time constant
$\Delta t$, then the average response over time can be expressed as [19]:

$$R(f_x, f_t) = \frac{1}{2\pi \cdot \Delta t} \cdot \left[ \frac{f_t}{f_t^2 + 1/(2\pi \cdot \Delta t)^2} \right] \cdot \left[ \sin(2\pi \cdot f_x \cdot \Delta x) \right] \cdot \tag{1}$$

The equation shows that the response depends on the spatial and temporal frequencies of the input and the
two parameters, $\Delta x$ and $\Delta t$. To see the effect of varying $\Delta t$, Fig. 2 shows contour plots of the responses for
three EMDs with different values of $\Delta t$ but the same $\Delta x$. We define the smallest time delay as a unit $\delta t$ and

the smallest distance as a unit $\delta x$. The y-axis represents $f_t$ in units of $\lambda_x/\delta t$ and the x-axis represents $f_x$ in units of $\lambda_x/\delta x$. Each set of concentric response contours represents the response of a single EMD with a particular $\Delta t$. The three EMDs have delays of $\Delta t = 1\delta t$, $2\delta t$ and $4\delta t$. For each EMD three contours and a peak are shown. Each contour represents a combination of spatial and temporal frequency of the input signal that produces the same average response. The peak average response at the center is 0.5 and the outermost contour represents a response of 0.4. This value has been chosen as a threshold to indicate that a detection is made only if the response is above that value. The slope of the line from the origin to any point on a contour represents the velocity, in units of $\delta x/\delta t$, of the input sinusoid corresponding to that point. If an array consists of EMDs with these three values of $\Delta t$ then the range of detectable velocities for the array would be given by the difference between the slopes of the lines that pass just above and below the outer contours of the three responses as shown on the plot. The total velocity range achievable with the three different values of $\Delta t$ is greater than for a single fixed $\Delta t$ but the spatial frequency range or bandwidth, shown on the x-axis, is the same as for any single $\Delta t$. An array of EMDs with different $\Delta t$'s can be arranged to cover a large range of velocities, however, the detectable spatial frequency range will not widen. Also, in practice it may be difficult to vary $\Delta t$ in an efficient and reliable way. In a standard imaging system $\Delta t$ would be constrained to be an integer multiple of the frame rate. In an integrated smart sensor implementation, the range for $\Delta t$ would be constrained by the size of the on-chip capacitors that would be needed for a low-pass filter.

To see the effect of varying $\Delta x$, Fig. 3 shows contour plots of the responses for three EMDs with different values of $\Delta x$ but the same $\Delta t$. The EMDs have input distances of $\Delta x = 1\delta x$, $2\delta x$ and $4\delta x$. Each EMD is sensitive to a specific range of velocities and spatial frequencies. An EMD with $\Delta x = 1\delta x$ is sensitive to lower velocities than an EMD of $2\delta x$. It is also sensitive to higher spatial frequencies. If an array of EMDs contains these three variations then both the velocity range and the spatial frequency range would be expanded. An array of EMDs with different values of $\Delta x$ can be designed to cover a wide range of velocities and spatial frequencies. Varying $\Delta x$ can be easily achieved either by varying photodetector spacing or, in the case of a standard imaging system, by grouping and averaging pixels into pixel aggregates, resulting in a space-variant imaging system.
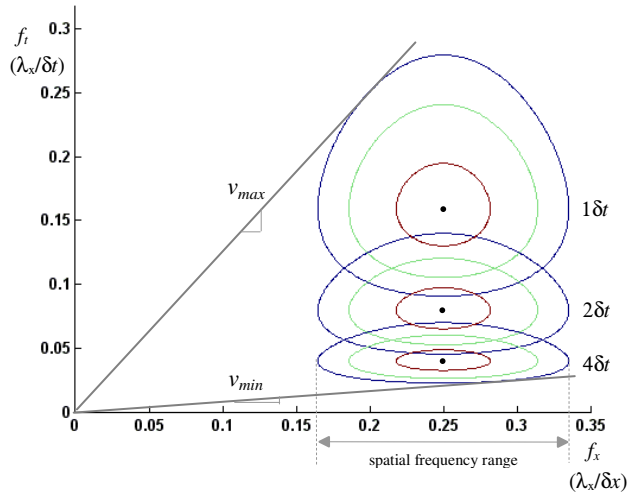
Fig. 2. Contour plot of the responses of three EMDs with delays of $\Delta t = 1\delta t$, $2\delta t$ and $4\delta t$. The range of detectable velocities for the three EMDs together is the difference in the slopes of the $v_{max}$ line and the $v_{min}$ line. The range of detectable spatial frequencies, shown on the x-axis, is the same as for a single EMD.
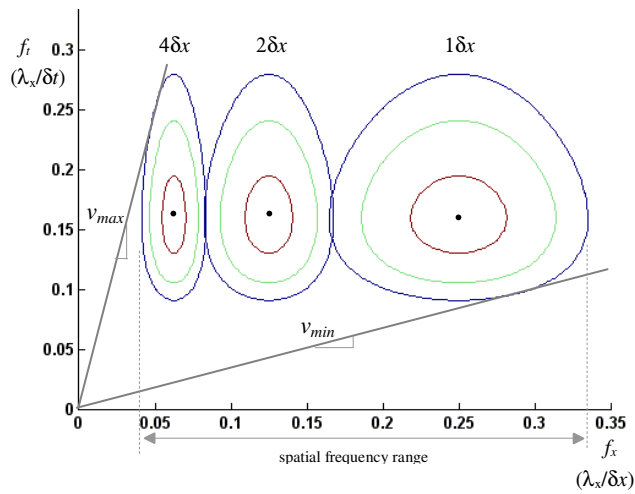


Fig. 3. Contour plot of the responses of three EMDs with input distances of $\Delta x = 1\delta x$, $2\delta x$ and $4\delta x$. The range of detectable velocities for the three EMDs together is the difference in the slopes of the $v_{max}$ line and the $v_{min}$ line while the range of detectable spatial frequencies, shown on the x-axis, is the combined ranges of the three EMDs.

## 2.2. Space-variant motion detection arrays and the EMD map

As mentioned previously, an array of EMDs with different values of $\Delta x$ can be easily achieved with a space-variant imaging system. A diagram of a very simple 1-D space-variant motion detection array is

shown in Fig. 4 (a). The array is arranged such that the EMDs with the smallest $\Delta x$ are at the center and the largest are at the periphery. The EMDs have values of $\Delta x$ that are in the same ratio of 1:2:4 as used in Fig. 3. Such a progression from the center to the periphery is similar in principle to that found in biological vision systems and is often referred to as a foveal array. Usually the region of highest resolution extends further out from the center to create a larger 'fovea' for high-resolution imaging. Many variations are possible and the diagram is meant only to show the idea of a space-variant motion detection array of tuned motion detectors. The ease with which $\Delta x$ can be varied in an implementation makes many different progressions possible. One detail that the diagram does show is that the image plane is not point sampled but that each sample extends to fill the space between samples.

To visualize the overall velocity range of an array like that of Fig. 4 (a) we can expand the diagram to show the velocity ranges of the individual EMDs as in Fig. 4 (b). The x-axis represents distance along the array in units of $\delta x$ and the y-axis represents the velocity range in units of $\delta x/\delta t$. The velocity can be either positive or negative and so each EMD is represented by a pair of rectangles. The width of each rectangle represents the width in the image plane occupied by the pair of photodetectors for that EMD in units of $\delta x$, and also provides an indication of the spatial frequency bandwidth for that EMD. The height of each rectangle represents the velocity range over which the EMD is able to detect motion. In an active vision system the array itself is moving to aim at the target so that velocities and positions are always relative to the velocity and position of the array. The y-axis is therefore labeled $v_r$ for relative velocity and the x-axis is labeled $x_r$ for relative position. As a target passes from left to right relative to the array we assume it has a positive relative velocity and if it passes from right to left it has a negative relative velocity. With this diagram, called the EMD map, we can now see that the periphery of the array covers higher relative velocities and lower spatial frequencies than the center of the array which covers lower relative velocities and higher spatial frequencies. This selectivity can help detect real-world targets with complex spatial frequency patterns. The lower spatial frequencies represent the gross overall pattern of the target while the higher spatial frequencies represent the fine details of the target and should only be resolved when the relative velocity between target and sensor is low [20].

Clearly the arrangement in Fig. 4 has an overall velocity range and spatial frequency range that would exceed the ranges for an array of uniform EMDs. The challenge with such a foveal array is to use the increased range of detection to control the movement of the array so that as the target becomes more aligned with the center of the array, its relative velocity will be lowered. It can then be detected with more precision by the smaller EMDs. The next section describes a model of a tracking system that incorporates a space-variant array of EMDs and that enables a method of analysis and design for the type of sensory-motor coordination needed for target tracking.
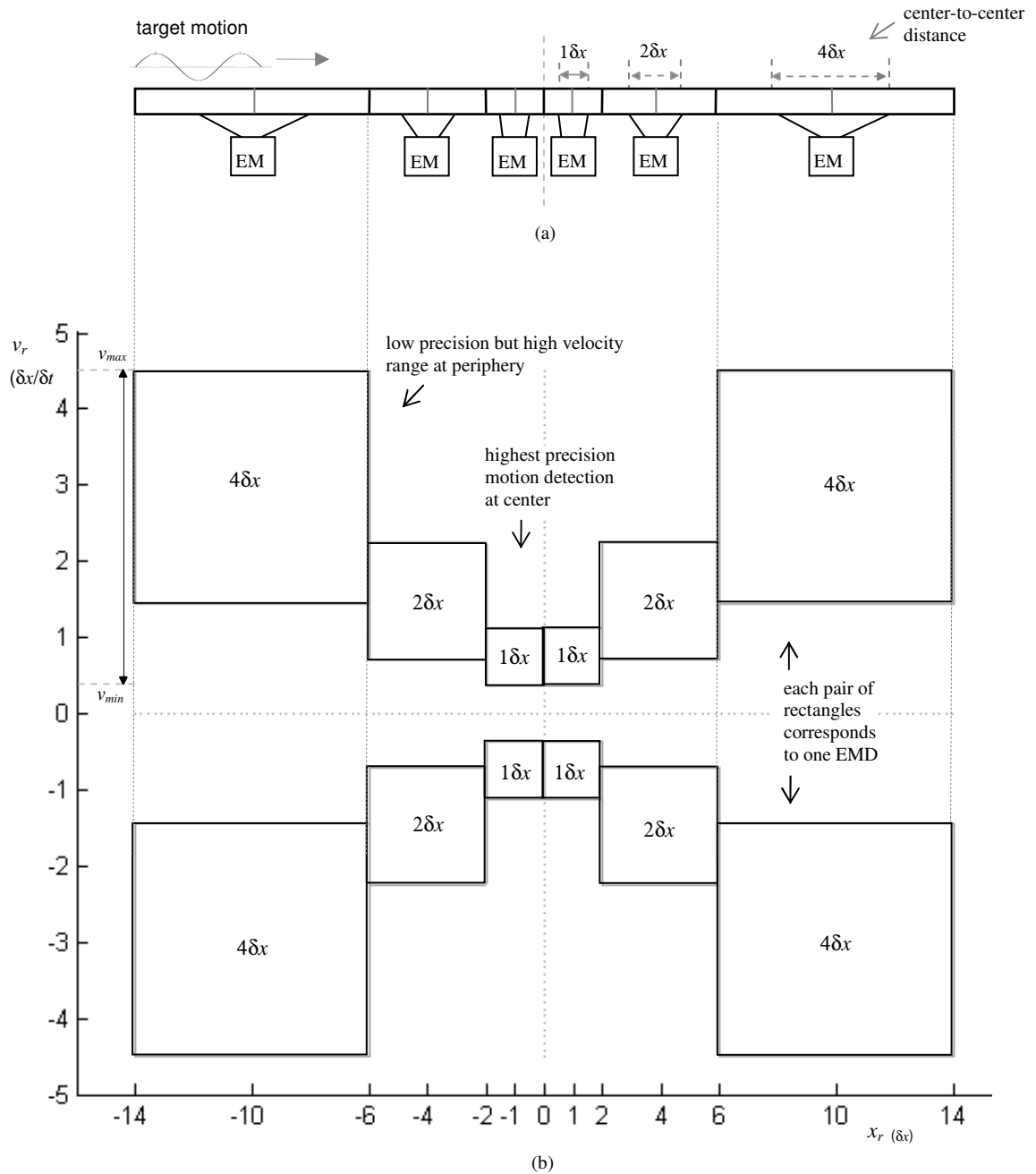
Fig. 4. A 1-D motion processing array (a) with an arrangement of EMDs symmetric about the center with $\delta x$ ratios of 1:2:4 and a plot (b) of the velocity ranges for the same 1-D array. Each rectangle and its mirror image about the x-axis represents a single EMD. The height of each rectangle represents the velocity range over which the EMD is able to detect motion.

## 3. System model for target tracking

A space-variant array of EMDs can be tuned to detect different velocity ranges and spatial frequencies. The detectable ranges can be visualized across the array with an EMD map. In this section we propose a method of incorporating such an array into an active vision system to make it possible to track targets over a wide range of velocities while still being able to achieve precise tracking. This section describes the

modeling, analysis and design of such a tracking system.

## 3.1. Modeling the tracking system

A diagram of a basic target tracking task is shown in Fig. 5. The target is assumed to be a rigid object moving in a plane at some fixed distance from a motion detection array which can pan left/right in order to keep its optical axis aligned with the target. For simplicity the motion detection array is assumed to be a 1-D array of EMDs like the one shown previously in Fig. 4(a). The target's position is $X_t(t)$ and its velocity is $V_t(t)$ in the same coordinate frame as the array's position $x_a(t)$ and velocity $v_a(t)$. When projected onto the array, the target's image position is $x_t(t)$ and its velocity is $v_t(t)$. Its appearance consists of a range of spatial frequencies that fall within the bandwidths of the different EMDs in the array. From this it can be assumed that a detection at an individual EMD will be made if any point on the target passes completely across the EMD from either direction within the velocity range of that EMD. It is possible that the target's image actually spans more than the distance between two EMDs so that multiple EMDs may simultaneously make a detection. For simplicity it is assumed that the response of one of these EMDs will be greater than the others and will be considered to be the only detection made.



**TARGET MOTION** $X_t(t)$ $V_t(t)$

**TARGET IMAGE MOTION** $x_t(t)$ $v_t(t)$
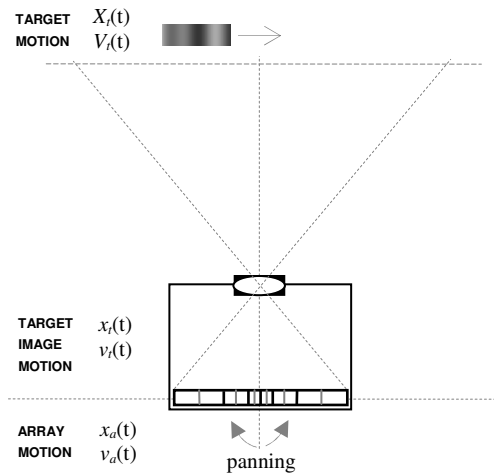
**ARRAY MOTION** $x_a(t)$ $v_a(t)$

panning

Fig. 5. Diagram of an active visual target tracking system based on a space-variant motion detection array.

As the target passes through the field-of-view of the array, detections will be made only when the velocity of the target relative to the array is within the range of one of the EMDs. Initially the relative velocity may be high and detections will only be possible at the periphery. Each detection results in a new estimate of the target's position and velocity relative to the array. This information can then be used to update the control signal used to drive the motion of the array. Through the controlled movement of the array, the relative velocity will be decreased and detections can be made closer to the center. Each new detection that is made is a *discrete event* that marks a change in the control signal. The arrangement of the

EMDs in the array acts as a switching mechanism that effectively 'switches' the system between different control signals depending on the particular detections that occur. The tracking performance of such a system depends on the interaction between the discrete events of detection and the continuous-time dynamics of the actuator/plant. This type of system is often referred to as a switched system or a hybrid control system [21][22].

Since detections and therefore 'switching' will occur at the boundaries of individual EMDs, an EMD map as shown in Fig. 4 (b) of section 2 describes those switching boundaries for a 1-D array. An indexing scheme for the map is needed to model the tracking system. Fig. 6 shows the same EMD map in which each rectangle is indexed with a quadrant number $q \in \{I,II,III,IV\}$ and an EMD number $n \in \{1, 2,...N\}$. As before, the y-axis represents the relative velocity $v_r$ and the x-axis represents the relative position $x_r$ of the target with respect to the array. Recall that the plot represents a 1-D array of EMDs so that quadrants I and II represent positive relative velocities while quadrants III and IV represent negative relative velocities for the same EMDs. When a detection is made the relative position and velocity of a target can be estimated resulting in a quantized measure of the actual values. The EMD map is essentially a non-uniform quantizer that partitions $x_r$-$v_r$ space into a non-exhaustive set of disjoint rectangles, each associated with an estimate of the relative position and velocity.



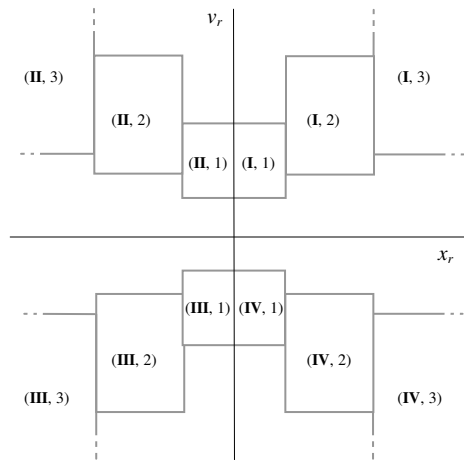Fig. 6. Indexing of the EMD map's rectangles or quantization regions in $v_r$-$x_r$ space showing quadrant numbers $q \in \{I,II,III,IV\}$ and EMD numbers $n \in \{1, 2,...N\}$.

The complete tracking system can be modeled as a basic closed-loop system with three components: the motion detection array, the controller and the actuator/plant. A block diagram of such a system is shown in Fig. 7. The target state vector is:

$$\mathbf{x}_t(t) = \begin{bmatrix} x_t(t) \\ v_t(t) \end{bmatrix} \qquad (2)$$

and the array state vector is:

$$\mathbf{x}_a(t) = \begin{bmatrix} x_a(t) \\ v_a(t) \end{bmatrix} \qquad (3)$$

These can be expressed in either angular or linear units, depending on the physical situation being modeled. The relative state vector is thus:

$$\mathbf{x}_r(t) = \mathbf{x}_t(t) - \mathbf{x}_a(t). \qquad (4)$$

When a detection is made at a particular $(q,n)$ in the EMD map, only an estimate of the relative state vector can be made:

$$\hat{\mathbf{x}}_r(q,n) = \begin{bmatrix} \hat{x}_r(q,n) \\ \hat{v}_r(q,n) \end{bmatrix}. \qquad (5)$$

A control signal $u(q,n)$ is then computed according to a control law. A very simple proportional control law computes $u(q,n)$ as the product of a gain vector $\mathbf{K}(q,n)$ and the estimate of the relative state vector, both of which depend on which EMD in the array has made a detection:

$$u(q,n) = \mathbf{K}(q,n) \cdot \hat{\mathbf{x}}_r(q,n) =$$
$$\begin{bmatrix} k_x(q,n) & k_v(q,n) \end{bmatrix} \cdot \begin{bmatrix} \hat{x}_r(q,n) \\ \hat{v}_r(q,n) \end{bmatrix} \qquad (6)$$

The value of the gain vector $\mathbf{K}(q,n)$ can be set differently for each rectangle in the EMD map. For each detection a new control signal will result from the value of the gain vector and a new estimate of the
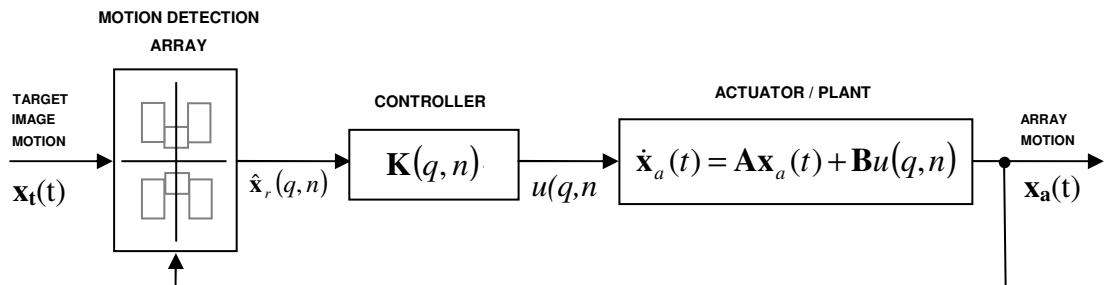


Fig. 7. Block diagram of a target tracking control loop with a space-variant motion detection array.

relative state vector. During the interval between detections the signal is unchanged resulting in a piecewise constant signal that drives the actuator/plant. The continuous-time dynamics of the actuator/plant can be described with a set of linear systems that are parameterized according to the quadrant and EMD number from the EMD map:

$$\dot{\mathbf{x}}_a(t) = \mathbf{A}\mathbf{x}_a(t) + \mathbf{B}u(q, n). \tag{7}$$

The state matrix $\mathbf{A}$ and the input matrix $\mathbf{B}$ depend on the dynamic characteristics of the actuator/plant. The combination of these continuous-time dynamics with the nonlinearity of the motion detection array makes for a difficult system to analyze.

*3.2. Phase-plane analysis and the sensory-motor map*

The analysis and design of such a system falls outside the scope of traditional control system approaches. Phase-plane methods and numerical analysis have mostly been used to study switched systems [22]. Some approaches to hybrid control systems with quantization in the loop [23] might be useful but the non-uniform and non-exhaustive nature of the EMD map makes analysis difficult. A way to visualize the interaction between the detection events and the continuous-time dynamics is needed to acquire insight into how the various parameters of the system will affect tracking performance. Using a simple example, this section describes such a method based on combining the EMD map with a phase-plane analysis of the system.

One way to visualize the tracking dynamics of the system in Fig. 7 is to plot the relative state vector, $\mathbf{x}_r(t)$, in the phase-plane. For a 1-D array we can plot relative velocity on the y-axis and relative position on the x-axis. In the case of successful tracking we expect both relative velocity and position to approach zero, likely in a cyclic trajectory about the origin. The exact trajectory would depend on the sequence of detections made by the EMDs in the array. An ordinary phase-plane plot, however, does not show the interaction between the detections made by the array and the changes in the trajectory. What is needed is a way to combine the information in the EMD map with the phase-plane trajectory in order to visualize the interaction so that we may design the system parameters for better tracking performance. To see how this can be done we will use a simple example of the system in Fig. 7. Consider a 1-degree-of-freedom tracking system that pans left/right with continuous-time dynamics described by a first-order system:

$$\begin{bmatrix} \dot{x}_a \\ \dot{v}_a \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -a \end{bmatrix} \begin{bmatrix} x_a \\ v_a \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix} u(q, n) \tag{8}$$

which has a single time-constant, $\tau = 1/a$. Such a system could serve as a simple model for a dc motor or for eye movement control. Assume that a 1-D array of EMDs with an EMD map, as shown previously in Fig. 4 (b), is being used. A good way to visualize the interaction between the detections that are made by the EMDs and the dynamics of (8) would be to transform the coordinates of (8), which could be radians and radians/sec., into the coordinates of the EMD map, which are $\delta x$ and $\delta x/\delta t$. By making this simple transformation it becomes possible to overlay a phase-plane plot of the relative state vector, $\mathbf{x}_r(t)$ onto the
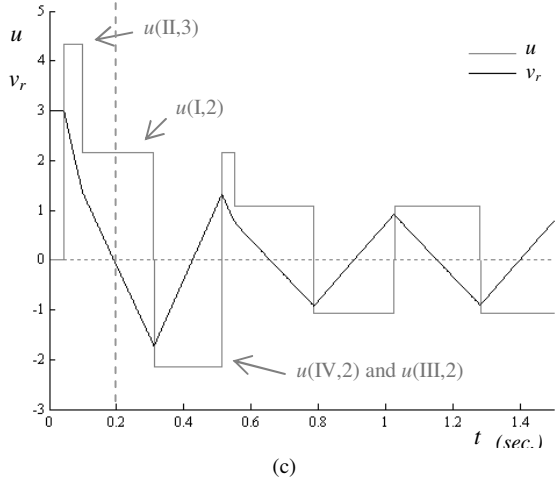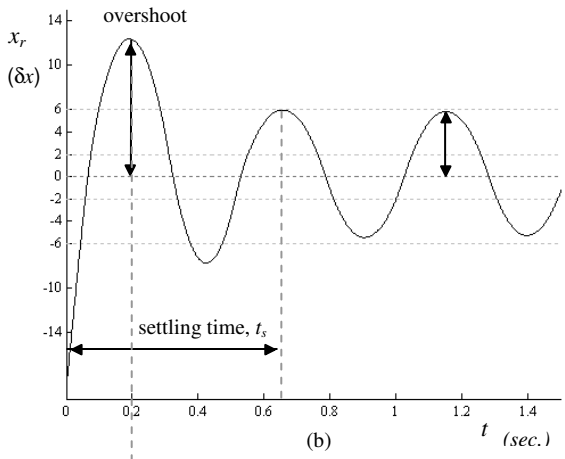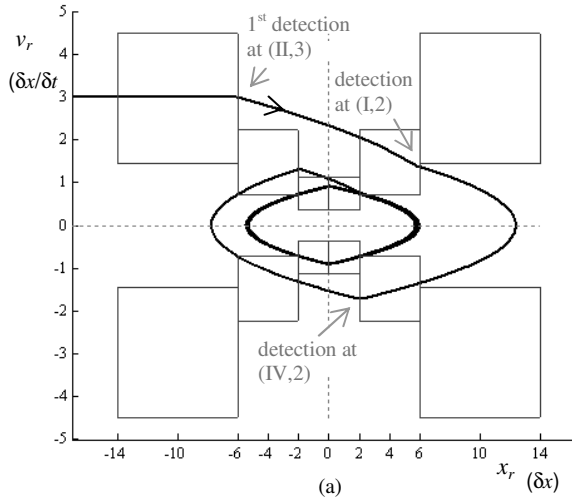
EMD map. Fig. 8 shows a Simulink block diagram that was used to compute and plot $\mathbf{x_r}$(t). The light grey triangles represent the transformation from rad./sec. to $\delta x/\delta t$ and the dark grey triangles represent the transformation from $\delta x/\delta t$ to rad./sec. The plot of the resulting trajectory in the phase-plane $x_r$-$v_r$ can then be superimposed onto the EMD map.



Fig.8. Block diagram for simulation of the tracking system with a tool like Matlab/Simulink. Light grey triangles represent transformation from rad./sec. to $\delta x/\delta t$. Dark grey triangles represent transformation from $\delta x/\delta t$ to rad./sec. Relative position and velocity ($x_r$, $v_r$) can be plotted on top of the system's EMD map.

For the particular example discussed above, Fig. 9 (a) shows a plot of the EMD map with a superimposed plot of the phase-plane trajectory of $\mathbf{x_r}$(t). The combination produces a sensory-motor map that reveals the interaction between the motion detection capabilities of the array and the tracking dynamics of the system. We assume that the motion detection array is initially stationary and a target is entering its field-of-view from the left at a speed of 3 $\delta x/\delta t$. The y-axis of the sensory-motor map corresponds with zero position error, the $x_r = 0$ line, and the x-axis corresponds with zero velocity error, the $v_r = 0$ line. The origin represents zero tracking error.

The sensory-motor map in Fig. 9 (a) shows ideal results obtained after manually tuning the value of the gain vector. To understand how the map can be used to visualize the interaction of the detection process with the tracking dynamics, it helps to follow the trajectory through the first three detections of the target. A detection is assumed to be made if the trajectory enters a rectangle on one side and leaves from the other side. When the first detection at (II,3) is made, one option is to compute a control signal to move the array to the left to reduce the position error. This would, however, raise the velocity error beyond anything that could be detected. Instead the correct reaction is for the array to move to the right to reduce the velocity error. The result of this motion is shown in the plot as the trajectory gradually slopes downwards towards zero velocity error. The trajectory eventually crosses the y-axis overshooting the $x_r = 0$ line. The second detection is made at (I,2), indicating that the target is moving toward the right edge of the field-of-view and so the array should continue to move to the right to decrease both position and velocity error. The trajectory eventually crosses the x-axis overshooting the $v_r = 0$ line. This means that the array is now moving strongly to the right to 'catch' the target and the relative velocity now becomes negative. When a detection is made

Fig. 9. Sensory-motor map (a) showing relative state trajectory $\mathbf{x_r}(t)$ superimposed onto EMD map along with relative position (b) and relative velocity and control signal (c) as functions of time.

at (IV,2) it means that the velocity error is now negative and the array should slow its rightward motion gradually so as to reduce velocity error. This is shown as the trajectory is pushed closer into the origin as it crosses the negative y-axis. This process continues until the trajectory enters a periodic cycle in which it orbits the origin at a distance that is limited by the size of the smallest EMDs.

The time response is plotted in Fig. 9 (b). It shows the correspondence between the shape of the trajectory and such time-domain metrics as overshoot, settling time and steady-state error. The settling time is the time needed for the trajectory to reach a periodic cycle and the steady-state error is given by the width of the cycle in the phase-plane. Time plots of the control signal, $u(q,n)$, and the relative velocity are shown in Fig. 9 (c).

By experimenting with the model it was found that achieving the type of trajectory in Fig. 9 (a) was only possible for an actuator/plant with a relatively large time constant. If the time constant $\tau = 1/a$ was too small then no amount of tuning of the gain vectors $\mathbf{K}(q,n)$ would lead to this trajectory. The trajectory would not cross the x and y-axes in a way that is needed for periodic orbiting. With a larger time constant and sufficiently large gain the array would be able to overshoot the target in a smooth and gradual way that is controllable with the gain. This is because the velocity of the plant should continue to increase/decrease between any two detections in the lower/upper half of the map. Its response to the control signal should be like that of an integrator. This means that the time constant should be several times larger than the time between any two detections. If $L$ is the total length of the array in pixels and $v_{rmin}$ is the smallest relative velocity of interest, in units of $\delta x/\delta t$, then as a guideline:

$$\tau \gg \frac{L}{v_{r\,min}} \cdot \Delta t \qquad (9)$$

For the array and trajectory of Fig. 9 (a), $L = 144\delta x$, $\Delta t = 0.01$ sec. and if we set $v_{rmin} = 0.5\ \delta x/\delta t$ then we need $\tau \gg 2.9$ sec. The actual value used for the plot was $a = 0.01$ or $\tau = 100$ sec. Fig. 10 shows a plot of $v_r$ and $u$ over time for $\tau = 0.1$ sec. The velocity of the array matches that of the target instead of continuing to decrease as shown by the dashed line. The result is that $v_r$ goes to zero asymptotically while $x_r$ increases without bound. A physical actuator/plant will have its own dynamics that may be very different from what is needed for good tracking performance. In that case, compensation can be used to reduce the effects of those dynamics and introduce a sufficiently large time constant into the system.
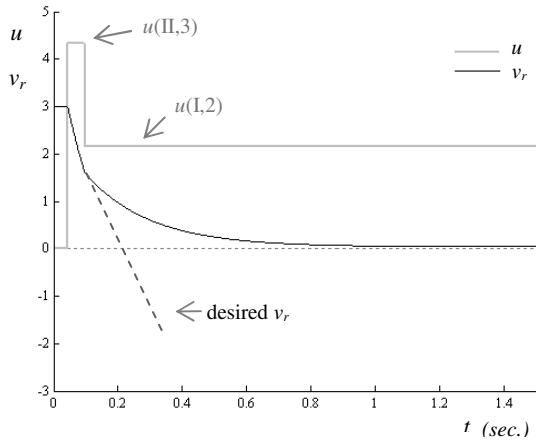
Fig.10. Plot of $v_r$ and $u$ over time for $\tau = 0.1$ sec. showing how the velocity of the array now matches that of the target instead of continuing to increase, as shown by the dashed line.

## 3.3. Designing with the sensory-motor map

The sensory-motor map makes it possible to explore the influence of the system parameters on the tracking performance by showing both the EMD map and the relative state trajectory together. From Fig. 9 (a) it can be observed that the array must be controlled to move in such a way that the trajectory will spiral inward gradually and enter into a cycle close to the smallest EMDs. As long as the relative velocity is positive the array should move to the right. When the relative velocity becomes negative then detections in quadrants III and IV should cause the array to slow its rightward motion. The control signal, $u(q,n)$, should be positive for detections in quadrants I and II and negative for detections in quadrants III and IV as shown in Fig. 9 (c). The vertical dashed line shows the correspondence between $v_r(t)= 0$ and the peaks of $x_r(t)$. The shape of the plot in (c) resembles the same type of plot for a simple nonlinear oscillator in which a relay with hysteresis and a linear system are connected in a feedback loop [24]. The relay is an odd function and therefore guarantees zero-mean cyclic signals. For our tracking system, $u(q,n)$ should also be an odd function symmetric about the x-axis. This is shown graphically in Fig. 11 (a). The values of $\mathbf{K}(q,n)$ must be set appropriately to achieve such a $u(q,n)$.

Looking back at Fig. 9 (a) it can be seen that a detection at (I,2) causes the trajectory to flatten out somewhat and thereby increase the overshoot. To reduce the overshoot it appears that all that is needed is to increase the control signal, and therefore $\mathbf{K}(q,n)$, for quadrants I and III as shown in Fig. 11 (b). Increasing $\mathbf{K}(q,n)$ in any case must be done incrementally however because if the slope of the trajectory becomes too sharp then it will not pass left/right through a rectangle which means that the target's motion is not detected. Without detections the trajectory will move outward away from the origin and the target will be lost. What is needed is a trajectory that spirals inward gradually but not so gradually that the settling time and overshoot are prohibitively large.
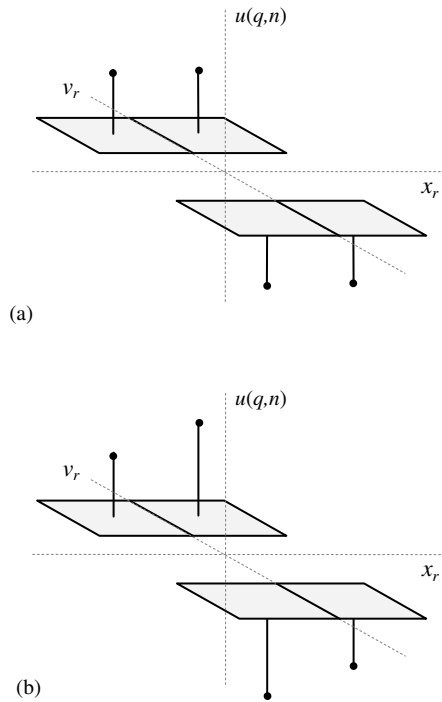
Fig. 11. (a) A simplified EMD map showing that $u(q,n)$ should be an odd function about the $x_r$ axis, resembling the action of a relay in a simple nonlinear oscillator. Only the rectangles for the smallest EMDs are shown to highlight the shape of the control signal. (b) Quadrants I and III are adjusted for less overshoot.

Insights provided by the sensory-motor map make it possible to develop a simple algorithm to determine values of $\mathbf{K}(q,n)$ that will lead to the desired settling time, steady-state error and overshoot for a particular combination of motion detection array and actuator/plant. Experiments with the simulation have shown us that large values of $\mathbf{K}(q,n)$ result in trajectories that slope sharply toward the origin, reducing the settling time. But if the slope is too sharp then detections will be missed and the target will be lost. The lower the relative velocity, the sharper the slope will be for a given $\mathbf{K}(q,n)$. Thus, if we consider a target entering the field-of-view of the array on the left or right, we should use the minimum relative velocity detectable by the outermost EMD, call it $v_{min}(q,N)$, when maximizing $\mathbf{K}(q,n)$. A simple algorithm would initialize $\mathbf{K}(q,n)$ to zero for all $(q,n)$ and iteratively increment all of the values by some step size. A good step size can be quickly found manually by trying a few values to see the effect on the shape of the trajectory.

A simple two-stage algorithm is used in which the first stage attempts to minimize settling time and steady-state error while the second stage attempts to minimize overshoot. For the first stage, $\mathbf{K}(q,n)$ is iteratively incremented causing the trajectory to slope more sharply toward the origin. Eventually the slope of the trajectory will be too sharp to be detected and the trajectory will move out of range. The values of $\mathbf{K}(q,n)$ for the previous iteration are those that will minimize the settling time and the steady-state error associated with that settling time. This simple procedure is shown as Stage 1 of the algorithm in Table I.

For higher relative velocities the trajectory will begin to have a larger value of overshoot. To reduce this overshoot, a second stage of optimization can be used. In this stage the values of $\mathbf{K}(q,n)$ for quadrants I and III only are iteratively incremented to increase the rate at which the trajectory will slope downward to cross the $x$-axis. The iterations continue until the trajectory suddenly diverges, indicating that a detection was missed. This is Stage 2 of the algorithm in Table I.

TABLE I
ALGORITHM TO SET $\mathbf{K}(q,n)$

Stage 1: Minimize settling time ($t_s$) and steady-state error $|x_r(t>t_s)|$

initialize: $\mathbf{K}(q,n) = 0$ for $\begin{cases} q = \text{I,II,III,IV} \\ n = 1, 2,...N \end{cases}$

do
· increment $\mathbf{K}(q,n) \leftarrow \mathbf{K}(q,n) + \text{step*}$ for $\begin{cases} q = \text{I,II,III,IV} \\ n = 1, 2,...N \end{cases}$
· solve system for $v_t(t) = v_{\min}(q,N)$
while $|x_r(t>t_s)| < |x_r(t>t_s)|_{\text{prev}}$

Stage 2: Minimize overshoot

do
· increment $\mathbf{K}(q,n) \leftarrow \mathbf{K}(q,n) + \text{step*}$ for $\begin{cases} q = \text{I,III} \\ n = 1, 2,...N \end{cases}$
· solve system for $v_t(t) = v_{\max}(q,N)$
while $|x_r(t>t_s)| < |x_r(t>t_s)|_{\text{prev}}$

* step size graphically determined

The algorithm is demonstrated with an example array of EMDs with values of $\Delta x$ in a progression of 1:2:3:4:5:6:7:8, resulting in the most overlap of velocity range for a system based on a standard camera. The plots in Fig. 12 (a)-(c) show the resulting sensory-motor diagrams after the first stage of the algorithm for target velocities of 4 $\delta x/\delta t$, 6$\delta x/\delta t$ and 8 $\delta x/\delta t$. As the target velocity is increased from (a) to (c) the trajectory approaches the origin less sharply and with more overshoot. The plots in Fig. 12 (d)-(f) show how the trajectories look after the second stage of adjustment for the same relative velocities as used in (a)-(c). The overshoot has been improved but the settling time has gotten worse. It is possible to continue in this way to tune the gain vectors individually to achieve the performance that a particular application calls for. More sophisticated algorithms to find optimal values of $\mathbf{K}(q,n)$ are possible and can easily be explored within this framework. The sensory-motor diagram is a visualization tool that helps in exploring the behavior of tracking systems that are based on space-variant motion detection arrays. It would be difficult to design such a tracking system without being able to make the observations that were made with the help of the diagram.
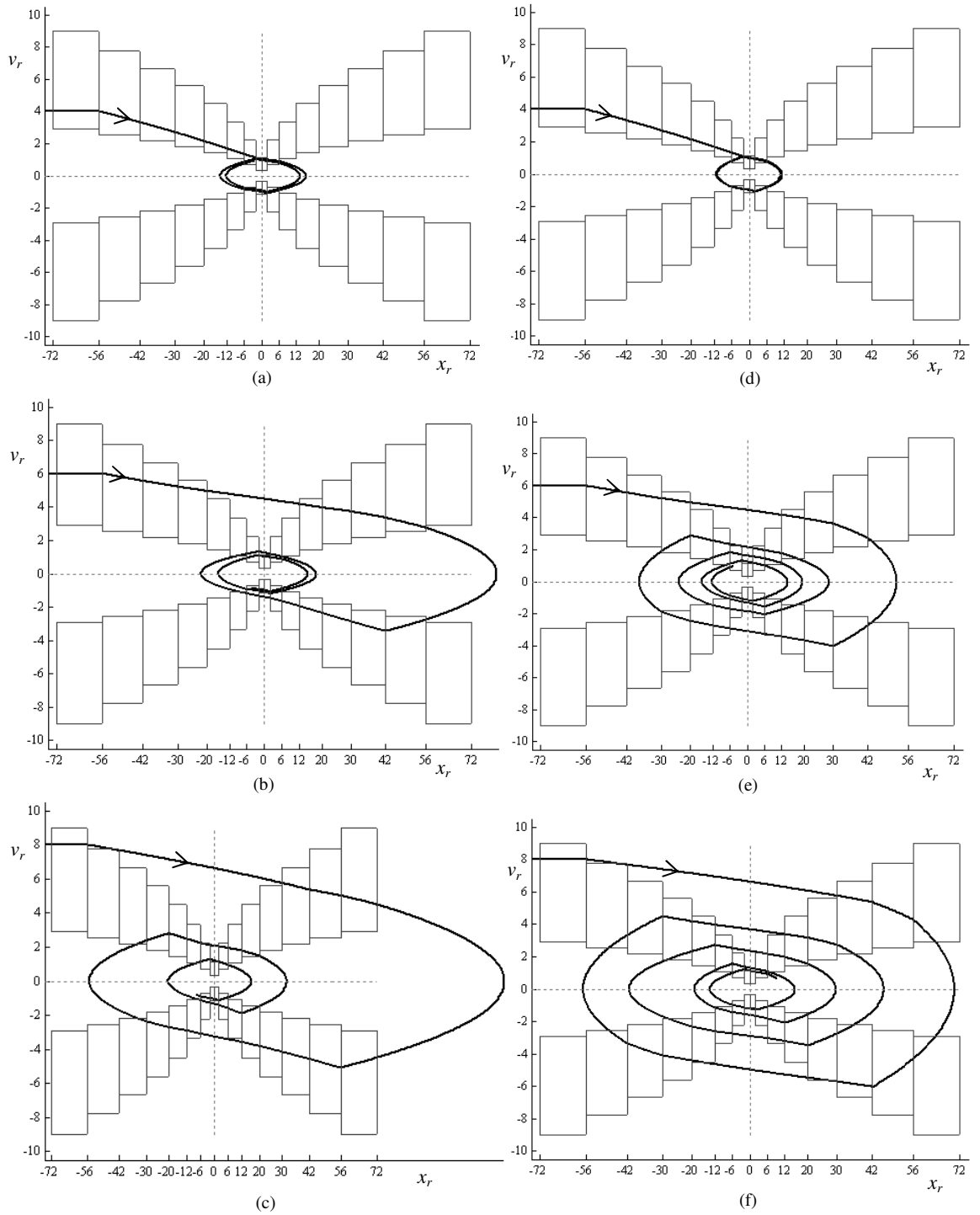
Fig. 12. Sensory-motor maps produced using the algorithm in Table I to set $\mathbf{K}(q,n)$. The first stage of the algorithm was used for (a)-(c) and the second stage was used for (d)-(f). The initial relative velocities are 4 $\delta x/\delta t$, 6$\delta x/\delta t$ and 8 $\delta x/\delta t$. The trajectories of (e) and (f) show much less overshoot but longer settling times than the corresponding trajectories in (b) and (c). For all plots $x_r$ is in units of $\delta x$ and $v_r$ is in units of $\delta x/\delta t$.

*3.4. Summary of the method*

An array of EMDs can be tuned to detect different velocity ranges and spatial frequencies by varying the distance, $\Delta x$, between the two points in the image plane where the inputs are acquired for each EMD. The detectable ranges can be visualized with an EMD map. The sensory-motor map makes it possible to design the tracking system by showing both the EMD map and the relative state trajectory together. The result is a system that combines both range and precision in a single array. Overall, the design of a space-variant tracking system involves two major steps. First, the EMD map must be defined to determine the distribution and tuning of the EMDs across the array. This is based on the needs of the application such as the range of velocities that must be detected and the minimum spatial feature that must be resolved and tracked. Second, a sensory-motor diagram of the tracking system should be used to plot the behavior of the tracking system for a range of target velocities. This can be done with any suitable numerical package such as MATLAB/Simulink. An optimization algorithm can then be used in conjunction with the plots to determine the control gain for each EMD.

## 4. Experiments

A sensory-motor system involves real-time interaction with a physical environment. Designing and implementing such a system is challenging because even in a structured setting the system must contend with various types of noise, vibration and the non-ideal characteristics of the electrical and mechanical components. To validate the design and evaluate the performance of the proposed approach in a physical setting, a complete active vision system was implemented and a set of target tracking experiments were performed.

*4.1. Experimental setup for tracking*

Fig. 13 (a) shows a schematic diagram of the active vision system and the experimental setup that was used to perform the target tracking experiments. Fig. 13 (b) shows a photograph of the setup. For all of the experiments the target is a rectangular paper card with a spatial frequency pattern printed on it. It is attached to a 2m long horizontal conveyor belt that is driven by a DC motor. A signal generator and amplifier are used to control the speed of the left-to-right movement of the target. Images of the moving target are captured by a standard video camera. The camera is mounted on a pan-head at a distance of $d =$ 0.8m from the target plane. The video output of the camera is connected to a desktop PC via a USB interface. A program written in C and the OpenCV library [25] processes the video in two stages. First, each frame's resolution is modified to create a foveated image according to a pre-specified EMD plot. Second, the output of each EMD operation is computed using the previous frame for the delayed inputs. This means the $\Delta t$ is fixed by the system's achievable frame rate, which in this case was approximately 15 frames per second. A separate thread in the program computes the control signal $u(q,n)$ at each frame interval and outputs a pulse-width modulated (PWM) signal to the RS232 serial port. That signal is then amplified in order to drive a DC motor on the pan-head to pan the camera horizontally.
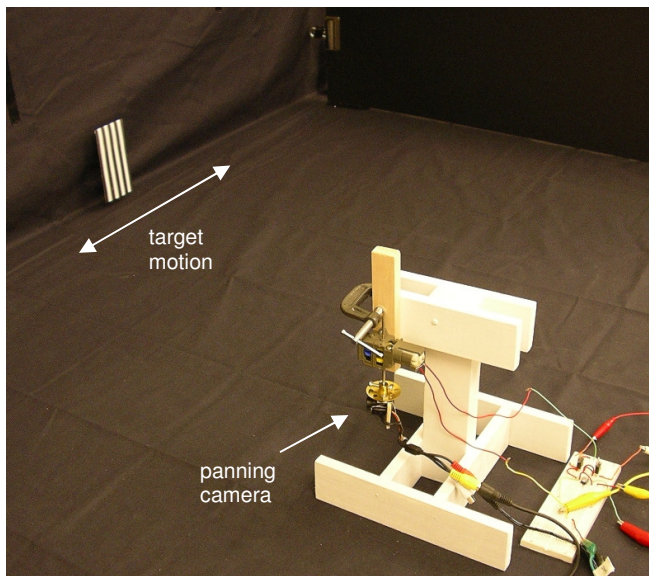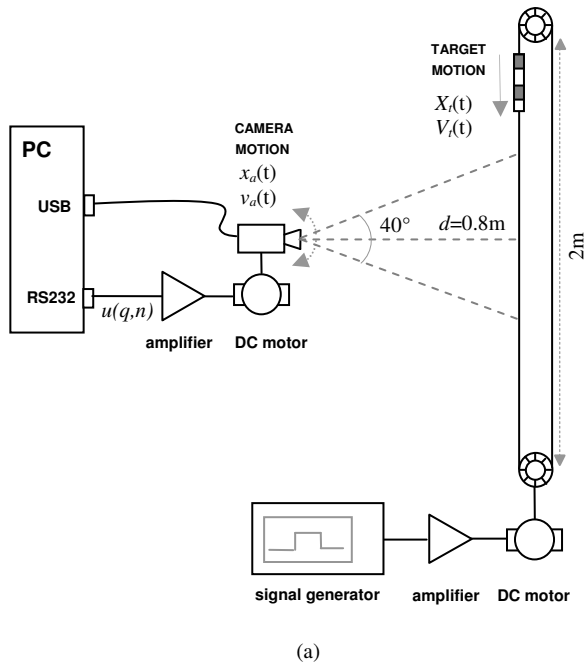
(a)



(b)

Fig. 13.  Schematic diagram (a) of experimental setup along with a
photograph (b) of the setup.

The video camera used in the experiments has a focal length of $f = $ 5mm. Its imaging area is
4.69mm×3.54mm and it captures images at a resolution of 640×480. This means that each pixel has a width
of approximately 7.3μm which is the $\delta x$ for this system. For each experiment the same EMD plot was used.
Fig. 14 shows a sensory-motor diagram for the experiment with a target initially moving to the right at 20
$\delta x/\delta t$. The algorithm of Table I was used to compute the values of $\mathbf{K}(q, n)$.

Assuming perspective projection, distances in the image plane are related to distances in the target plane by:

$$x_t \frac{d}{f} = X_t \tag{10}$$

Speeds in the image plane are related to speeds in the target plane by:

$$1 \frac{\delta x}{\delta t} \cdot \frac{d}{f} = 0.0175 \, m/s \tag{11}$$

in which $\delta x = 7.3 \times 10^{-6}$ m and $\delta t = 1/15$ sec. Of course $d$ actually changes with time because the camera is panning but the target is translating. This alters the target's image size and speed slightly but the approximation of a constant $d$ was found to be acceptable when the camera only pans through an angle of about 90°. Equation (11) can be used to calculate the speed of the target's motion needed for the experiments. For example, 20 $\delta x/\delta t$ in the image plane is 35cm/s in the target plane, which can be achieved by adjusting the signal generator that drives the target's conveyor belt.
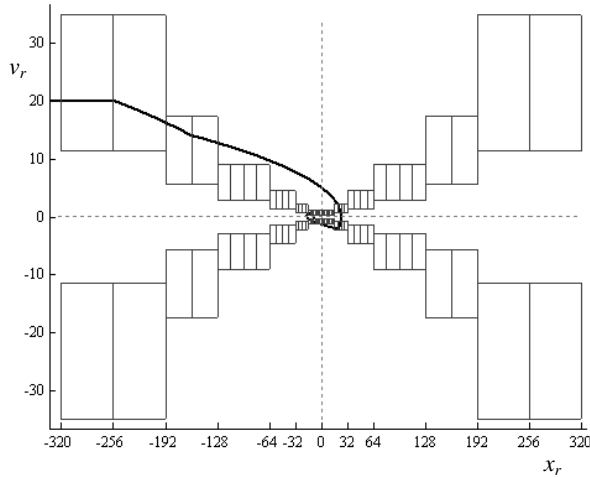


Fig. 14. Sensory-motor map for the experimental system.

*4.2. Experiments and results*

Each experimental run consisted of initially positioning the target at the left extreme of the conveyor belt and pointing the camera such that the target was just outside the left edge of its field-of-view. The target was then driven to the right at one of three pre-selected speeds. Fig. 15 shows the tracking results for speeds of (a) 20 $\delta x/\delta t$, (b) 25 $\delta x/\delta t$ and (c) 30 $\delta x/\delta t$. Each plot shows the estimate of the relative position, $x_r$, as a function of time. The values were measured by the program for each frame processed.

For each run the relative position converged to the expected oscillation about the zero error axis as predicted by the model. As expected, the settling time, $t_s$, increased with target speed as shown in the plots. The overshoot was also expected to increase with target speed but the results in Fig. 15 (a) and (b) show, with dashed circles, times at which the tracking system reacted too quickly and then produced overshoots when compensating for the overreaction. This occurs twice in (a) likely due to the slow speed of the target at 20 $\delta x/\delta t$, but only once for (b) when the target was moving at 25 $\delta x/\delta t$. Because of the way the system was implemented it was difficult to control the time constant of the tracking dynamics. The use of a software generated PWM signal and the PC's serial port undoubtedly led to unpredictable delays in the feedback loop and made it difficult to control the system's performance precisely.

Despite the minor discrepancies between the experimental results and the model, the experiments showed that even a crude implementation of the proposed approach enables target tracking over a wide range of speeds while still being able to achieve precise tracking. It should be emphasized that no special calibration of the camera or of the experimental setup was required. The camera and platform could be oriented in a very approximate way without significantly affecting the results.
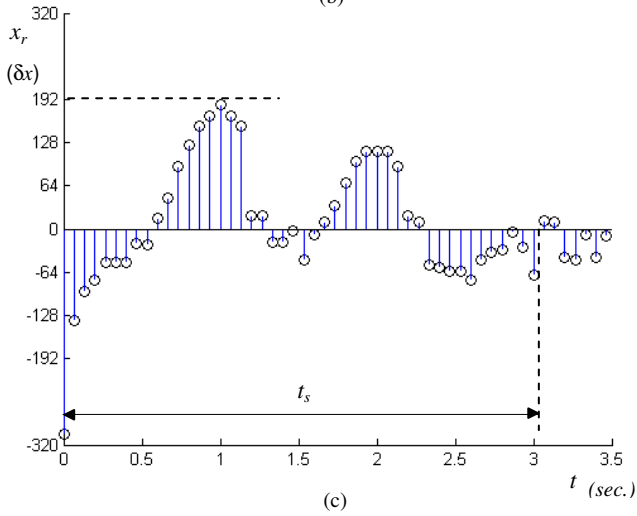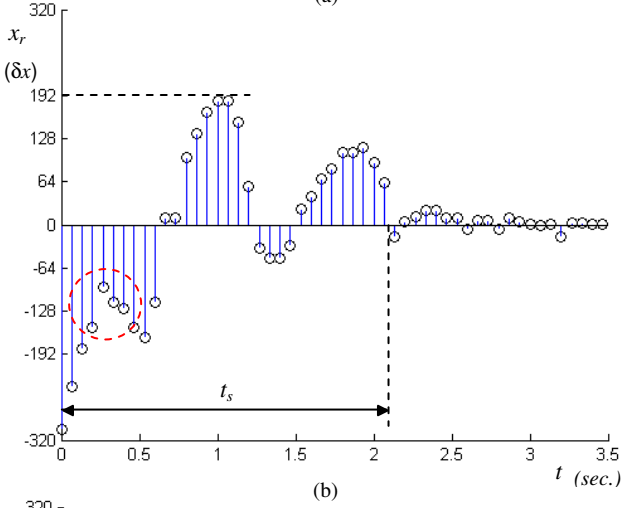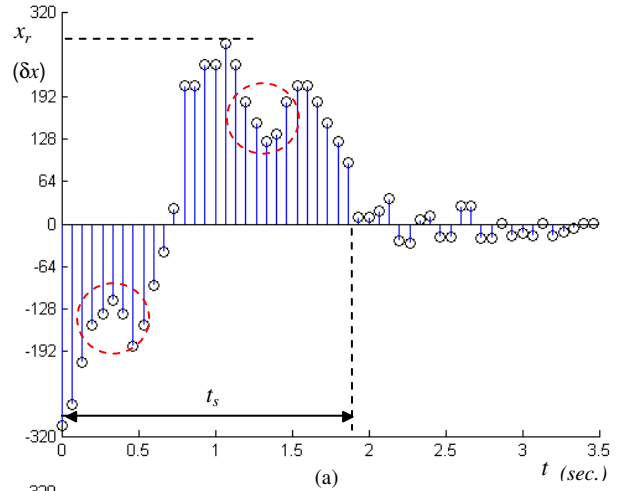
Fig. 15. Tracking results for target speeds of (a) 20 $\delta x/\delta t$, (b) 25 $\delta x/\delta t$ and (c) 30 $\delta x/\delta t$. The horizontal dashed lines indicate overshoot and the vertical dashed lines indicate settling time.

## 5. Conclusion

A space-variant motion detection system has been proposed for the task of active visual target tracking with the goal of tracking targets both precisely and over a wide range of speeds in a computationally effiecient manner. The system makes use of three strategies observed in biological systems: space-variant or foveal sensing, a spatio-temporal frequency based model of motion detection and the alignment of sensory-motor maps. Space-variant imaging is used to create an array of EMDs that can be used to track targets over a wide range of speeds while still being able to precisely track fine motion. A method of analysis and design of the tracking system has also been proposed. Based on a switched-system model it uses a sensory-motor map which provides a graphical link between the imaging geometry and the dynamic performance of the tracking task. Several simulations show how the method can be used to control tracking performance such as overshoot and settling time. The resulting system exhibits both range and precision in a computationally efficient manner. To validate the method, experiments with a real active vision system have been performed.

There are several ways in which the approach could be improved and expanded upon. We have only worked with 1-D arrays. The sensory-motor diagram uses two dimensions (velocity and distance) to describe a one-dimensional motion detection array. An extension of the technique to a two-dimensional motion detection array will depend on the geometry of the array. If the array is arranged radially with concentric circles of detectors, it may be possible to create a three-dimensional representation by having a sensory motor diagram for each angle of detectors in the array. This could lead to interesting three-dimensional maps that may provide greater insight into target tracking and other applications. One such application is obstacle avoidance, in which the desired trajectory would move away from the origin rather than towards it. We have also limited ourselves to a single target on a black background. It would be interesting to explore how the approach could be used to improve figure-ground separation in real-world scenes.

Overall the approach represents an effort toward a more general treatment of sensory-motor systems. It may be possible to use the proposed sensory-motor map in other sensory-motor contexts such as in haptic devices and auditory systems. Being able to visualize both sensing and motor actions on the same plot makes it possible to design effective systems that require fewer computational resources. When the sensing is well matched with the motor activities, and both are well-matched with the task, then the entire system can operate more efficiently. In addition, this approach may help researchers to model real biological systems and gain insight into biological mechanisms of sensory-motor control.

## References

[1] R. Wallace, P.W. Ong, B. Bederson, E.L. Schwartz, "Space variant image processing," *International Journal of Computer Vision*, 13(1): 71–90, 1994.

[2] G. Sandini and V. Tagliasco, "An anthropomorphic retina-like structure for scene analysis," *CGIP*, vol.14:365-- 372, 1980.

[3] 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.

[4] M. Tistarelli, and G. Sandini, "On the advantages of polar ad log-polar mapping for direct estimation of time-to-impact from optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 401-410, 1993.

[5] C.F.R. Weiman, R. Juday, "Tracking algorithms using log-polar mapped image co-ordinates," *Intelligent Robots and Computer Vision VIII: SPIE Proceedings on Intelligent Robots and Computer Vision*, pp. 843-853, 1998.

[6] J. Boluda and J. Domingo, "On the advantages of combining differential algorithms and log-polar vision for detection of self-motion from a mobile robot," Robotics and Autonomous Systems**,** Volume 37, Issue 4, 31 December 2001, Pages 283-296

[7] E. L. Schwartz, "Anatomical and physiological correlates of visual computation from striate to infero-temporal cortex," *IEEE Trans. Syst., Man,Cybern.*, vol. SMC-14, pp. 257–270, Feb. 1984.

[8] E. C. Hildreth and C. Koch, "*The analysis of visual motion: From computational theory to neuronal mechanisms*," Ann. Rev. Neurosci., vol. 10, pp. 477--533, 1987.

[9] C. Fennema and W. Thompson, "Velocity determination in scenes containing several moving objects," *Computer Graphics and Image Processing* 9, pp. 301—315, 1979.

[10] B. K. P. Horn and B. G. Schunk. "Determing optical flow," *Artificial Intelligence*, 17:185--203, 1981.

[11] E. H. Adelson and J.R. Bergen, "Spatiotemporal energy models for the perception of motion," *Journal of the Optical Society of America A*., vol. 2, no. 2, pp. 284--299, Feb 1985.

[12] W. Reichardt,**,** "Autocorrelation, a principle for the evaluation of sensory information by the central nervous system," in Sensory Communication, ed. W.A. Rosenblith, pp. 303-317, MIT Press, Cambridge, Massachusetts, 1961.

[13] Schrater, P. R. and Simoncelli E. P. (1998) Local Velocity Representation: Evidence from Motion Adaptation. Vision Research. 38(24), pp. 3899-3912, 1998.

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

[15] J. Kramer, R. Sarpeshkar, and C. Koch, "Pulse-based analog VLSI velocity sensors," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 86-101, 1997.

[16] C.M. Higgins, and S.A. Shams, "A Biologically-Inspired Modular VLSI System for Visual Measurement of Self-Motion," *IEEE Sensors Journal*, pp. 508--528, December 2002.

[17] F. Ruffier and N. Franceschini, "Optic flow regulation: the key to aircraft automatic guidance," Robotics and Autonomous Systems**,** Volume 50, Issue 4, 31 March 2005, Pages 177-194

[18] Stein, B.E. and Meredith, M.A., *The Merging of the Senses*, Cambridge, MA: The MIT Press, 1993.

[19] R. O. Dror, D. C. O'Carroll, and S. B. Laughlin, "Accuracy of velocity estimation by Reichardt correlators," *Journal of the Optical Society of America A* 18, pp. 241–252, February 2001.

[20] D. Claveau and C. Wang, "A space-variant sensor structure for complex target detection and tracking," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006 Proceedings*, Volume: 3, Page(s): III-896- III-899, 21-24 May 2006.

[21] A. van der Schaft and H. Schumacher, *An Introduction to Hybrid Dynamical Systems*, Springer-Verlag, London, 2000.

[22] D. Liberzon, *Switching in Systems and Control*, Birkhauser, Boston, 2003.

[23] R. W. Brockett and D. Liberzon. "Quantized feedback stabilization of linear systems," *IEEE Transactions on Automatic Control*, 45:1279--1289, July 2000.

[24] W. Richard Kolk and Robert A. Lerman, *Nonlinear System Dynamics*, New York : Van Nostrand Reinhold, 1992.

[25] Open Source Computer Vision Library, http://www.intel.com/technology/computing/opencv/