Proc. SPIE Intelligent Robots and Computer Vision XII, Boston, MA, September, 1993.

Predictive visual tracking

Albert J. Wavering and Ronald Lumia

Robot Systems Division
National Institute of Standards and Technology
Gaithersburg, MD 20899

ABSTRACT

Due to delays in image acquisition and processing, prediction is a critical factor for successful visual tracking of moving objects (both for humans and for vision machines). This paper explores some alternative techniques for predicting object motion for the purpose of tracking with an active camera system. In particular, one of our goals is to develop a system that will track an object undergoing "random" motion quite well, but that will track much better (at higher speeds with less lag) if the object settles into a periodic motion of some kind. Rather than identify parameters for specific signal models to accomplish this, we propose to use a finite set of previous samples of the target signal for the signal model. The advantages and problems associated with this approach are discussed. Results of experiments using different prediction algorithms with TRICLOPS, a high-performance active vision system, are also presented.

1. INTRODUCTION

Image acquisition and processing delays are present in all animal and machine vision systems. If no effort is made to compensate for these delays, visual tracking performance is severely limited. The difficulties of image processing delays in visual servoing systems and proposals for mechanisms which intend to alleviate the problem have been presented by a number of researchers.[2,6,7,15,16,19] Any visual tracking system which attempts to make the system perform as if the image processing delay did not exist must incorporate a mechanism for predicting the motion of the target. Typically this involves the use of a predictive filter, such as one of the constant-coefficient Kalman filters for kinematic models (α–β and α–β–γ filters[4,11]), along with some assumptions about the target motion (e.g., it will be smooth and limited in acceleration). A similar technique involves performing a least-mean-square fit [LMSF] of a polynomial function of time to a sequence of previous target positions.[17] The LMSF polynomial may then be extrapolated to obtain estimates of future target positions. Both the constant-coefficient Kalman and polynomial LMSF predictive filters perform acceptably for low frequencies, but exhibit problems such as overshoot when the target motion contains higher frequencies. A quite different approach, called Target-Selective Adaptive Control (TSAC) is proposed by Bahill and McDonald[2,15] for tracking predictable targets without latency. In this approach, the waveform of the target motion is compared with a number of possible candidates of periodic motions. If there is a match, parameters of the target waveform are identified from the target motion signal, and the resulting signal model is used to perform zero-latency tracking.

The purpose of this paper is twofold. First, we will examine the capabilities and limitations of "conventional" (fixed-parameter) predictive filters by looking at their predicting performance in the frequency domain. Second, we will present a hybrid system in which low frequency motions are tracked using a conventional predictive filter, while higher frequency periodic motions are tracked using a TSAC-inspired approach. However, instead of using predefined waveform types, our proposed approach for tracking periodic motions will use previous samples of the target signal itself as the model for target motion, and autocorrelation will be used to determine when the target motion is predictable. With this approach, there is more flexibility in the tracking of different target motion waveforms.

1.1. BACKGROUND

Before proceeding, it will be helpful to provide some background information regarding the experimental visual tracking hardware and general tracking algorithm to be used with the prediction techniques discussed in this paper. The predictive tracking algorithms which will be discussed have been developed for use with a high-performance active vision system which was designed and built at the National Institute of Standards and Technology. The system, called TRICLOPS (The Real-time Intelligently-ControLled Optical Positioning System),[9,14,21] employs a direct-drive design to achieve high-bandwidth position servo control and dynamic performance comparable to the human eye. A photo of TRICLOPS is shown in Fig. 1. TRICLOPS is a four degree-of-freedom device, and it has three cameras; a center wide-angle field-of-view camera and a pair of outboard high-resolution *vergence* cameras. In the tracking algorithms discussed in this paper, only the outer vergence cameras are used. The system is controlled with a VME-based multiprocessing system. Additional details regarding the design, performance, and control and image processing hardware and software of TRICLOPS may be found in Fiala et al,[9] Lumia,[14] and Wavering et al.[21]

A block diagram of the general approach to visual tracking is given in Fig. 2. The visual tracking described in this paper

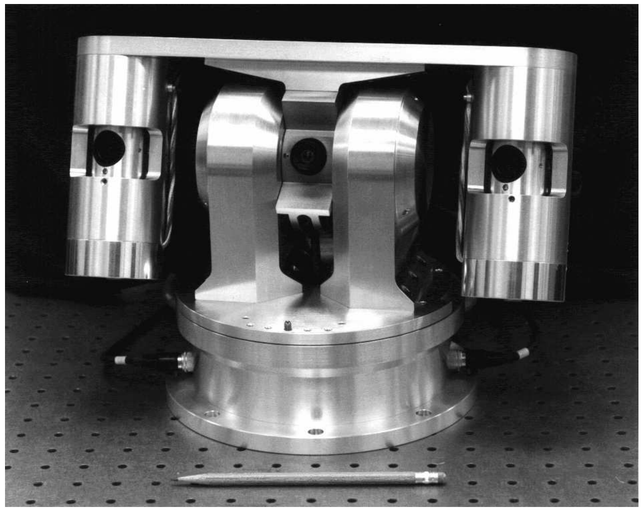

Fig. 1. TRICLOPS active vision system.

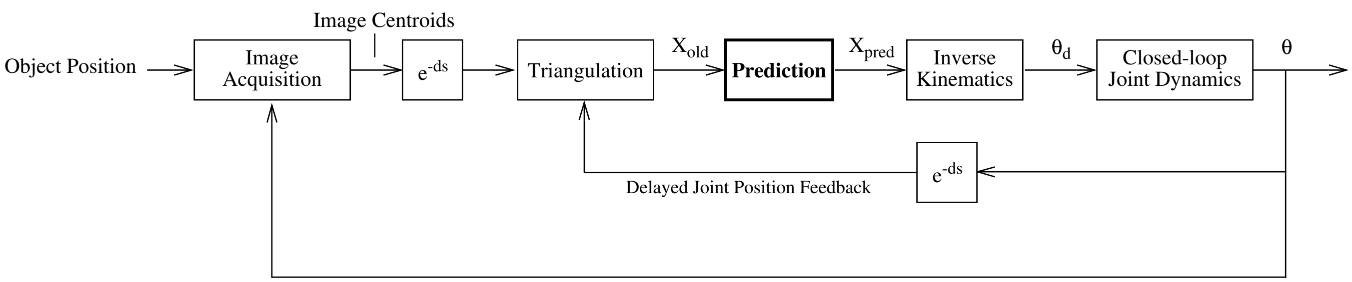

Fig. 2. Block diagram of tracking algorithm..

uses triangulation to determine the position of the target. To reduce the image processing delay, this example application uses simple thresholding to locate a high-contrast target in each camera image. Image centroids are obtained at a rate of 30 Hz, and they have an *approximate* image acquisition and processing latency of 0.084 s (i.e., in the figure d = 0.084). In order to compute the target position with respect to a fixed (with respect to the base of TRICLOPS) coordinate system, delayed values of the joint position feedback which correspond in time to the centroid data are required. After triangulation, a delayed estimate of the target position is available (X_{old}). This value is used as the input to a predictive filter, the performance of which is the focus of the current paper. The joint angles required to point the cameras at the predicted target position are then determined using inverse kinematics. The joint goal positions, which are updated at 30 Hz, are linearly interpolated into subcommands to the joint servos which are updated at a rate of 2 kHz. This interpolation means that it will take approximately 0.033 s to execute the motion to the new goal position, so the prediction should provide a total of approximately 0.084 + 0.033 = 0.117 s. The tracking approach described above has also been advocated by Murray et al.[16]

As with the system proposed by Robinson[19] to explain human oculomotor behavior, the effect of the inherent negative visual feedback is cancelled in the above approach to create an effectively open-loop system with regard to the visual determination of the object position. The inclusion of positive feedback of the delayed joint positions accomplishes this,[10] which (if the estimate of image processing delay is accurate) effectively decouples the estimate of the target position from the motion of the cameras. Instead of relying on an open-loop plant to execute the desired pointing motions, however, a high-bandwidth joint position controller is used. This is a special case of the Smith predictor[20] approach proposed by Brown et al[6,8] and Bahill.[3] However, in these previous systems (and the classic Smith predictor) a model of the controlled plant is required since the delay in the system is assumed to be distributed between the controller and the plant. Because the time response of the joint servos of TRICLOPS is very small compared with the visual delays in the system, a model of the controlled plant is not needed—the actual joint positions obtained from the real plant are used instead.

1.2. IMPORTANCE OF ACCURATE DELAY DETERMINATION

It is extremely important to recognize the sensitivity of the system to inaccuracies in the estimate of the visual delay to be used in the joint position feedforward path (triangulation). If the delay estimate is accurate, then the measurement of the target position is minimally affected by motions of the cameras, and it does not matter if the cameras are pointing directly at the target or not, as long as the target is visible to the cameras. In this case, the system is effectively open-loop with respect to the visual information, and the high-bandwidth joint servos track the stream of goal positions as if the target position estimates were coming from an external measurement system. As a result, the positive phase introduced by the predictive filter will not cause the system to become unstable. In fact, if the delay estimate is accurate enough, it is possible to maintain stability even if *greater* prediction than is necessary is applied—in which case the system will *lead* the target. However, any errors in the delay estimate applied to the joint positions will result in coupling between the joint motions and the predicted position. A feedback loop is created in which the feedback loop gain is equal to the delay error times the velocity[7]. If the gain of this loop is negative (caused by underestimating the delay), then the effect is primarily increased tracking error. However, if the visual delay is overestimated, a positive feedback loop is created which can cause instability. Our observations (both in simulations and on the real system) have indicated that overestimation will result in instability if the full amount of prediction is used. However, if the prediction is decreased and the delay error is small, stability can often still be achieved. The necessity of reducing prediction to maintain stability has also been noted by other authors.[8]

Since accurate estimation of the visual delay is so critical to the stability of the system, we have recently implemented a mechanism by which time stamps are used to measure the delay for each sample, rather than relying on an average empirically-determined or calculated value. This measured delay is then used to retrieve the corresponding delayed joint values (to a resolution of 0.5 ms) from a queue of previous joint positions maintained in a common memory communications buffer. Using this mechanism for delay compensation, the system does in fact remain stable even if more than the required amount of prediction is applied.

Given the use of high-bandwidth joint position servos and accurate determination of the target position using delayed position feedback, performance of the visual tracking system is determined predominately by the adequacy of the predictions supplied by the predictive filter. To get an idea of the frequency response characteristics of the tracking system, then, we can examine the magnitude and effective prediction (phase) of the predictive filter in isolation. The next section will provide such an analysis for two classes of predictive filters; constant coefficient Kalman filters (α–β and α–β–γ filters) and polynomial least-mean-square fit (LMSF) filters. A discussion of how the frequency response of the tracking system can be extended for predictable signals by adding correlation-based tracking is also included. The performance of the prediction techniques when implemented on the experimental hardware is presented in Section 3.

2. FREQUENCY RESPONSE CHARACTERISTICS OF PREDICTIVE FILTERS

Although predictive filters have been widely used in visual servoing and target tracking applications,[1,4,8] relatively little attention has been given to their frequency response characteristics. This is particularly true for the case where the filter is actually used to predict, rather than to just smooth. The analyses presented in this section provide insight to the two most salient difficulties of predictive filters: 1) magnitude overshoot at high frequencies and 2) frequency-dependent prediction time.

2.1. The α–β and α–β–γ Filters

The α–β and α–β–γ filters are constant-coefficient Kalman filters which are intended to be used to model and predict target motion based on assumptions of constant velocity and acceleration (respectively) between sampling intervals. Using the notation of Blair et al[5] the filter equations for the α–β filter are given by the following:

Prediction

$$x_p(k) = x_s(k-1) + T v_s(k-1), \quad v_p(k) = v_s(k-1) \tag{1}$$

Smoothing

$$x_s(k) = x_p(k) + \alpha[x_m(k) - x_p(k)], \quad v_s(k) = v_p(k) + \frac{\beta}{T}[x_m(k) - x_p(k)] \tag{2}$$

where

$x_s(k)$ = smoothed position at time k	$x_p(k)$ = predicted position at time k
$v_s(k)$ = smoothed velocity at time k	$v_p(k)$ = predicted velocity at time k
$x_m(k)$ = measured position at time k	T = time period between measurements,
α = filter coefficient for position	β = filter coefficient for velocity.

To use the filter to predict the position x_f at some future time τ, the following equation is used:

$$x_f = x_s(k) + \tau v_s(k). \tag{3}$$

Three dimensional predictive filtering is accomplished using a separate filter for each dimension.

The optimal relationship between α and β in (2) is given[11] by

$$\beta = 2(2-\alpha) - 4\sqrt{1-\alpha}. \tag{4}$$

For the α–β–γ filter, the state equations are
Prediction

$$x_p(k) = x_s(k-1) + Tv_s(k-1) + \frac{T^2}{2}a_s(k-1), \quad v_p(k) = v_s(k-1) + Ta_s(k-1), \quad a_p(k) = a_s(k-1). \tag{5}$$

Smoothing

$$x_s(k) = x_p(k) + \alpha[x_m(k) - x_p(k)], \quad v_s(k) = v_p(k) + \frac{\beta}{T}[x_m(k) - x_p(k)], \tag{6}$$

$$a_s(k) = a_p(k) + \frac{\gamma}{2T^2}[x_m(k) - x_p(k)]. \tag{7}$$

where

$a_s(k)$ = smoothed acceleration at time k $\quad$ $a_p(k)$ = predicted acceleration at time k
γ = filter coefficient for acceleration.

To use the α–β–γ filter to predict the position x_f at some future time τ, the following equation is used:

$$x_f(k) = x_s(k) + \tau v_s(k) + \frac{1}{2}\tau^2 a_s(k). \tag{8}$$

The optimal value of γ is given[11] by

$$\gamma = \frac{\beta^2}{\alpha} \tag{9}$$

To examine the frequency response of the α–β and α–β–γ filters, the filter transfer functions must be determined. To do this, the state equations for each filter were represented in the form of a signal flow graph, and then the transfer function was determined using Mason's general gain rule (as given in reference [12]). This is the same procedure used in Blair et al[5] to determine the frequency of a two-stage α–β–γ filter. The resulting transfer functions are

$$\frac{X_f(z)}{X_m(z)} = \frac{\left(\alpha + \frac{\beta\tau}{T}\right) + \left(\beta - \alpha - \frac{\beta\tau}{T}\right)z^{-1}}{1 + (\alpha + \beta - 2)z^{-1} + (1-\alpha)z^{-2}} \tag{10}$$

for the α–β filter and

$$\frac{X_f(z)}{X_m(z)} = \frac{\left(\alpha + \frac{\beta\tau}{T} + \frac{\gamma\tau^2}{4T^2}\right) + \left(\beta - 2\alpha + \frac{\gamma}{4} - \frac{2\beta\tau}{T} - \frac{\gamma\tau^2}{2T^2} + \frac{\gamma\tau}{2T}\right)z^{-1} + \left(\alpha - \beta + \frac{\gamma}{4} + \frac{\beta\tau}{T} + \frac{\gamma\tau^2}{4T^2} - \frac{\gamma\tau}{2T}\right)z^{-2}}{1 + \left(\alpha + \beta + \frac{\gamma}{4} - 3\right)z^{-1} + \left(\frac{\gamma}{4} - 2\alpha - \beta + 3\right)z^{-2} + (\alpha - 1)z^{-3}} \tag{11}$$

for the α–β–γ filter. The α–β and α–β–γ filters are seen to be infinite impulse response (IIR[18]) filters.

Given the filter transfer functions, the magnitude and phase response can be computed for different combinations of filter gains (α) and prediction amounts (τ) using standard discrete Fourier transform (DFT/FFT) techniques.[18] The resulting magnitude and phase curves for several different combinations of parameters are given in Figures 3-6. The frequency scale for these

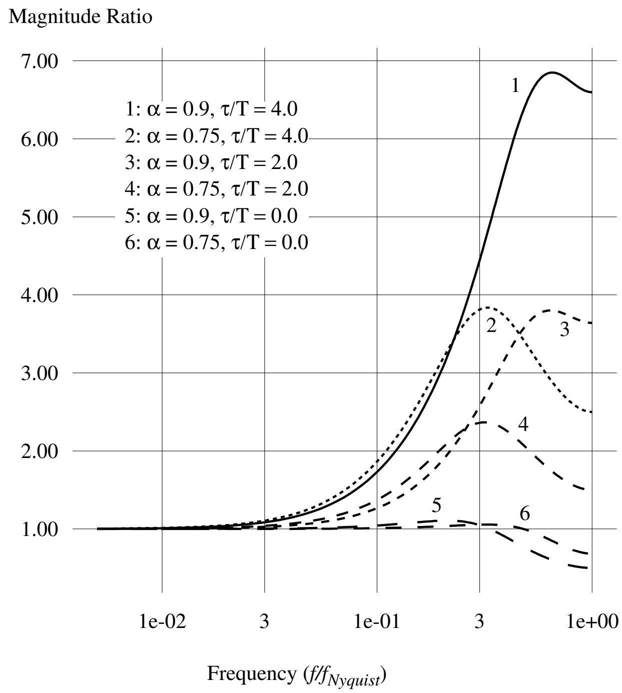

Fig. 3. Magnitude response of α–β filter.

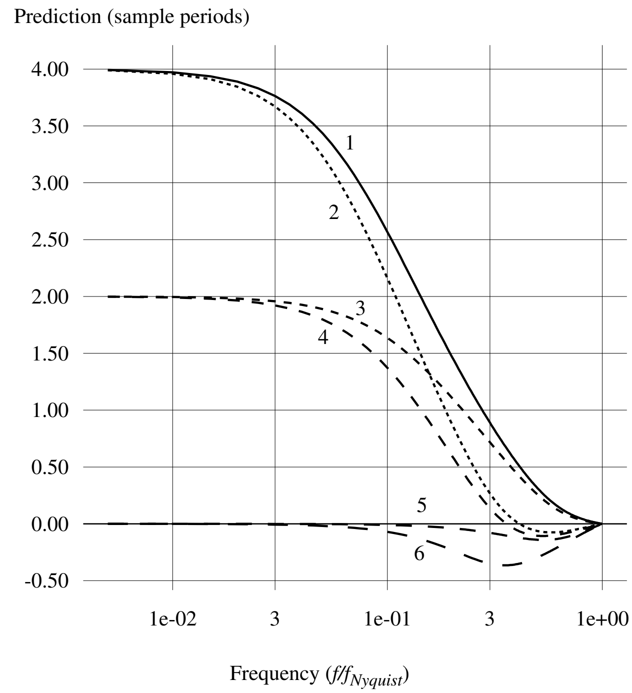

Fig. 4. Prediction characteristics of α–β filter.

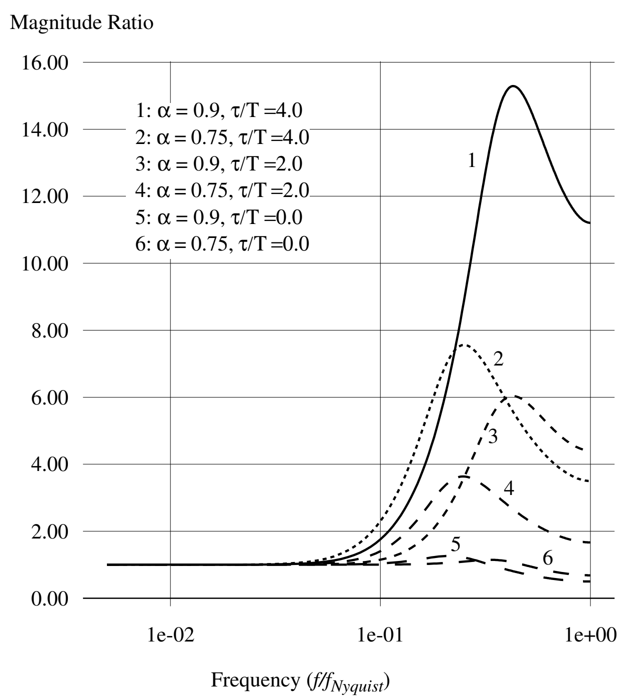

Fig. 5. Magnitude response of α–β–γ filter.

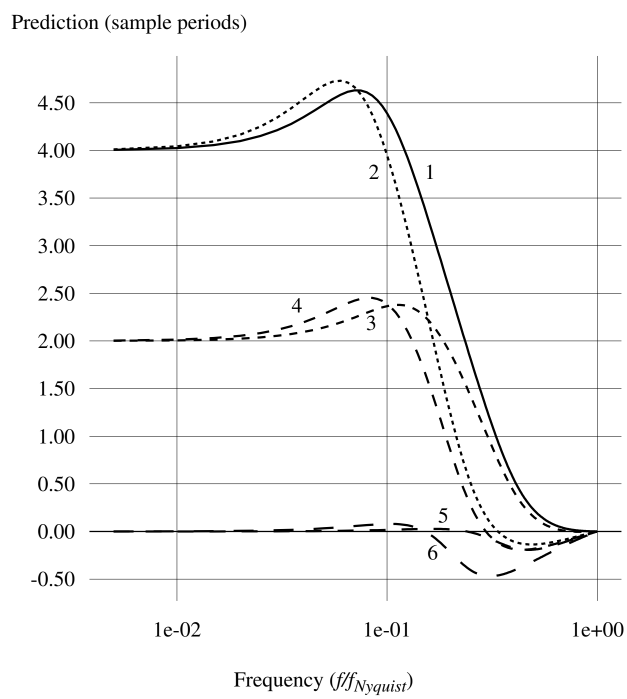

Fig. 6. Prediction characteristics of α–β–γ filter.

curves is in terms of fractions of the Nyquist frequency[13]

$$f_{Nyquist} = \frac{1}{2T}. \tag{12}$$

For each value along the frequency axis in Fig. 4 and Fig. 6 the phase response has been plotted in terms of effective prediction using the relationship

$$\tau_{eff}/T = \frac{\phi}{2\pi fT} = \frac{\phi}{\pi(f/f_{Nyquist})} \tag{13}$$

where

τ_{eff}/T = effective prediction in sample periods
ϕ = phase in radians and
f = frequency in Hz.

One of the most prominent characteristics that emerges from these plots is the peak in the magnitude ratio which occurs for all combinations of parameters for both the α–β and the α–β–γ filters, even when the amount of prediction is zero (in this case the filters just smooth the data). The height of the magnitude ratio peak is closely related to how much prediction is being performed. Therefore, large amounts of prediction can be expected to result in much overshoot at higher frequencies. High peaks in the amplitude response also indicate increased sensitivity to noise in that frequency range, as well. The higher-order α–β–γ filter has a higher peak in the amplitude response when compared with an α–β filter using the same gains, but it also gives a wider range of flat frequency response.

Increasing the filter gains results in a somewhat wider range of flat magnitude response, but it also increases the sensitivity to measurement noise. This is consistent with the notion of the *tracking index*,[11] which is proportional to the ratio of the position maneuverability, or signal, uncertainty to the position measurement uncertainty. The tracking index is central to the development of the optimal gain relationships (4) and (8). The optimal value of α increases as the tracking index increases, in which case the uncertainty in the target motion becomes relatively more significant than the position measurement uncertainty. If the target maneuverability and measurement noise characteristics are known, they may be used to compute the optimal Kalman filter gains directly. In practical applications of these filters for visual tracking, however, the parameter α is typically adjusted empirically to find the largest value which will give acceptable performance in the presence of the measurement noise.

Since these filters are intended for use in prediction, there is often a tacit assumption that they have constant prediction vs. frequency characteristics. As indicated in Fig. 4 and Fig. 6, however, this is not the case. The phase characteristics of the predictive filters result in prediction performance which is frequency-dependent. The α–β filter provides the desired amount of prediction at low frequencies, but the amount of prediction rolls off starting at about 1-3% of $f_{Nyquist}$. The α–β–γ filter also starts off with the correct prediction, and rolls off at higher frequencies, but in the transition region there is an increase in the effective prediction up to a peak value. The height of the prediction peak is determined primarily by the nominal amount of prediction, and secondarily by the value of α. This prediction peak can be beneficial in that less than full prediction can suffice at low frequencies (the target will be tracked with some lag but will remain in the field-of-view) and then as the frequency increases, tracking will become more accurate. Using reduced prediction is beneficial in terms of reducing the sensitivity to noise.

2.2. Polynomial LMSF Filters

Another straightforward approach to predictive filtering is to perform a least-squares fit of a finite-length vector of previous position data to a polynomial function of time. When used to filter data (without performing any prediction), such a procedure is sometimes referred to as a *least-mean-square-fit (LMSF) smoother*.[17] Let

$$\boldsymbol{x}_m = H\boldsymbol{b} + \boldsymbol{w} \tag{14}$$

where

$\boldsymbol{x}_m = \left[x_m(k)\ x_m(k-1)\ \dots\ x_m(k-N+1)\right]'$ = vector of N samples of measured positions

$\boldsymbol{b} = \left[b_0\ b_1\ \dots\ b_M\right]'$ = vector of polynomial coefficients

$\boldsymbol{w} = \left[w_0\ w_1\ \dots\ w_{N-1}\right]'$ = vector of residual errors ($'$ denotes the transpose operation) and the $N \times (M+1)$ matrix H is given by

$$H = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & -T & (-T)^2 & \ldots & (-T)^M \\ . & . & . & \ldots & . \\ 1 & (1-N)T & ((1-N)T)^2 & \ldots & ((1-N)T)^M \end{bmatrix}$$ with T = the sample time as before.

The least-squares solution to the determination of the polynomial coefficients[17] is

$$\boldsymbol{b} = (H'H)^{-1}H'\boldsymbol{x}_m. \tag{15}$$

When the filter is used to predict a future position at time τ, the output equation for the filter is

$$x_f = [1\tau^2\tau^3 \ldots \tau^M]'\boldsymbol{b} = [1\tau^2\tau^3 \ldots \tau^M]'(H'H)^{-1}H'\boldsymbol{x}_m. \tag{16}$$

From (16) it is clear that the polynomial LMSF equations may be written in the z-domain as

$$X_f(z) = c_1 x_m + c_2 x_m z^{-1} + c_3 x_m z^{-2} \ldots + c_N x_m z^{-(N-1)},$$

which is the form of a finite impulse response (FIR) filter[18]. The vector of filter coefficients is given by

$$\boldsymbol{c}' = [1\tau^2\tau^3 \ldots \tau^M]'(H'H)^{-1}H'. \tag{17}$$

As an example, the filter equation for a quadratic polynomial ($M = 2$), fitted to $N = 6$ position samples, with sample time $T = 1.0$ s, and predicting $\tau = 2.0$ s into the future is

$$x_f(z) = 2.3571x_m + 0.2429x_m z^{-1} - 0.9429x_m z^{-2} - 1.2000x_m z^{-3} - 0.5286x_m z^{-4} + 1.0714x_m z^{-5}. \tag{18}$$

As with the $\alpha-\beta$ and $\alpha-\beta-\gamma$ filters before, the magnitude and prediction response of the polynomial LMSF predictive filters may be computed using the DFT. Figures 7-10 give the response characteristics for quadratic and cubic polynomial filters for different combinations of N and prediction time (in terms of the number of sample periods, τ/T). These plots show that the LMSF filters share many of the same general characteristics of the $\alpha-\beta$ and $\alpha-\beta-\gamma$ filters. There is a peak in the magnitude ratio response which increases with the amount of prediction (Fig. 7, Fig. 9). The prediction response also exhibits a peak in the transition region (Fig. 8, Fig. 10). As with the gain parameter α, the parameter N in the LMSF filter is chosen to provide a compromise between being able to track highly uncertain targets and being able to perform well in the midst of measurement uncertainties. The parameter N is selected by decreasing the value to the smallest which will give acceptable performance given the amount of measurement noise present. Increasing the order of the filter extends the flat response band, but results in higher magnitude peaks and greater noise sensitivity. Specific comparisons between the constant-coefficient Kalman and LMSF filter types are somewhat difficult to make, because the filter parameters α and N are not directly comparable. However, some generalizations can be made. The magnitude response peaks of the LMSF filters are higher, but narrower than those which occur with the $\alpha-\beta$ and $\alpha-\beta-\gamma$ filters. The peak in the LMSF filter prediction vs. frequency curve is quite a bit larger than that for the $\alpha-\beta$ and $\alpha-\beta-\gamma$ filters. The frequency range for flat magnitude response of the LMSF filters tends to be wider than that obtained with the $\alpha-\beta$ and $\alpha-\beta-\gamma$ filters.

2.3. Prediction of Periodic Motions Using Autocorrelation

The polynomial LMSF and $\alpha-\beta-\gamma$ filters do a good job of predicting signals which are have relatively low frequency content. These techniques may be the best one can do for signals which are unpredictable, but smoothly-changing. However, if the signal is truly predictable (periodic, or nearly so), then it should be possible to identify the periodic waveform and use the data of the previous cycle(s) to track at higher frequencies without latency or overshoot. This is the general idea behind the Target-Selective Adaptive Controller (TSAC) proposed by Bahill and McDonald.[2,15]

In the TSAC approach, two methods are suggested for adaptively computing the signal to track. The first requires the identification of the target waveform from a menu of possible selections (triangular, sinusoidal, parabolic, etc.), and the subsequent automatic determination of waveform parameters such as target amplitude and frequency. The second method uses a second-order difference equation to compute the synthesized signal. It is desirable to be able to adapt to and predict signals in such a

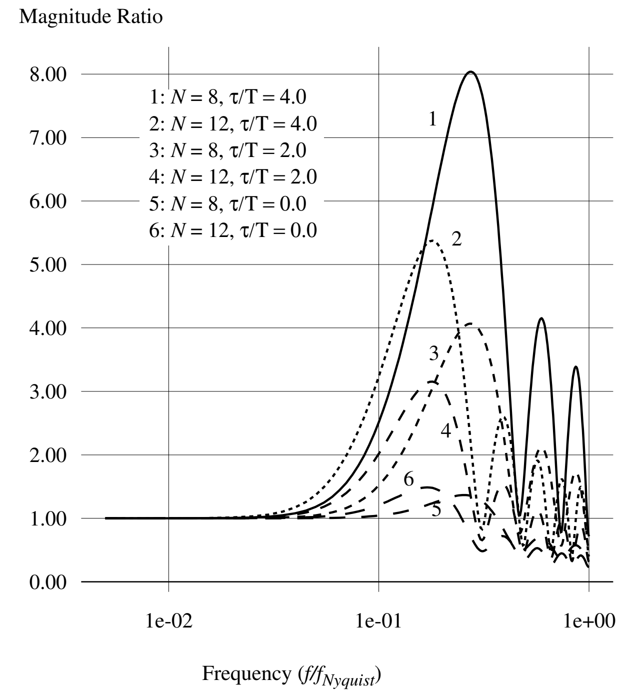

Fig. 7. Magnitude response of quadratic LMSF filter.

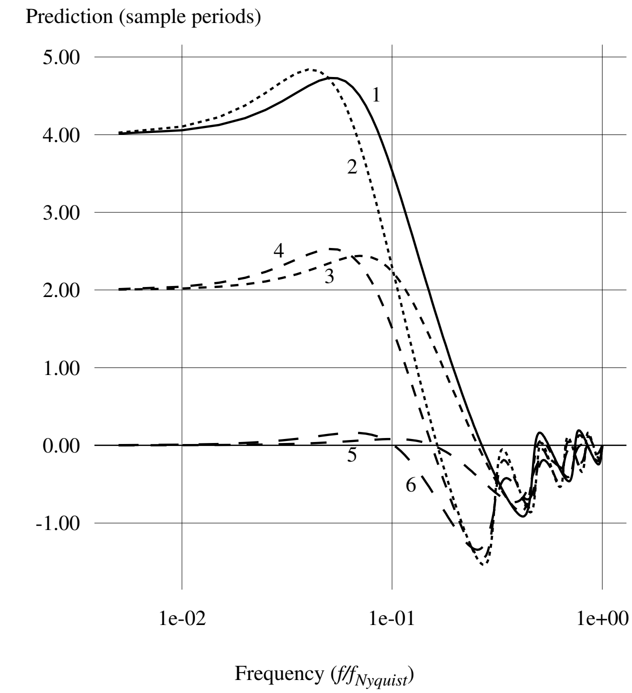

Fig. 8. Prediction characteristics of quadratic LMSF filter.

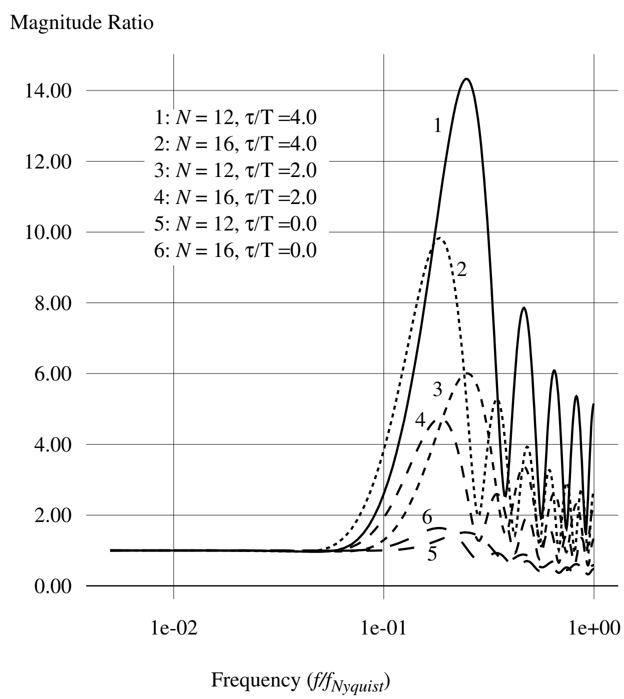

Fig. 9. Magnitude response of cubic LMSF filter.

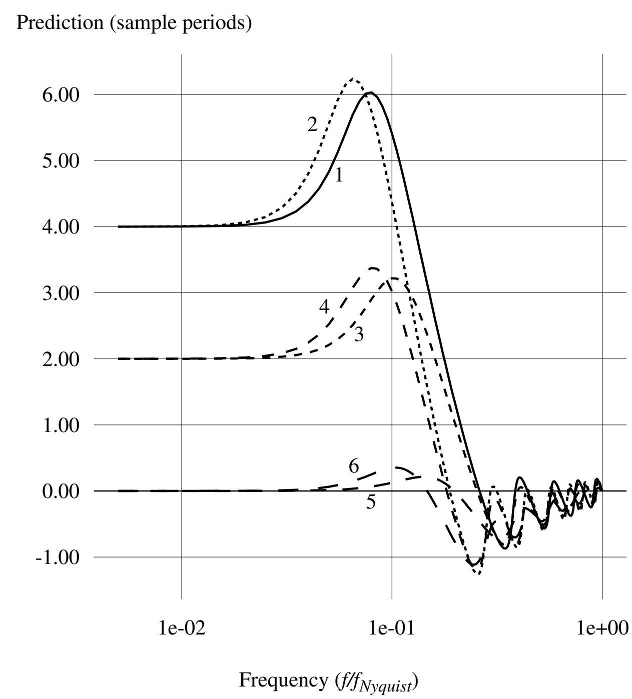

Fig. 10. Prediction characteristics of cubic LMSF filter.

way that the structure of the signal does not need to be identified a priori. One way to do this is to compute an autocorrelation function on the target signal. In this technique, a window of sampled data is incrementally shifted back in time and compared with previous sample windows. For each value of time delay, a correlation function is computed between the two sample windows. A strong correlation indicates a periodic target signal. This is analogous to looking for stereo matches in left and right camera images. If there is a good match, then the "future" state from the previous occurrence can be used to predict the future state for the current situation. If the signal follows the previous pattern, then the prediction will be made without overshoot or latency.

An implementation of this type of prediction has been developed. To determine the correlation with time-shifted values of past positions, the sum of the squared errors between corresponding samples in the two windows is computed, as given below:

$$E_D = \sum_{k=0}^{N-1} (x(k) - x(k+D))^2. \tag{19}$$

where

E_D = the sum of the squared errors for a time shift of D sample periods,
$x(k)$ = k^{th} sample in sample window (with $x(0)$ = the most recent sample),
D = number of sample periods of time shift, and
N = number of samples in window interval.

Using this correlation for prediction of periodic motions involves the following steps:

1. Compute E_D for integer values of D in the range $N \geq D \geq 1$
2. Identify the minimum E_D
3. If E_D is less than a threshold value, then use $x(D- N_\tau)$ for the predicted value (where N_τ = the number of cycles to predict ahead). Otherwise, use one of the conventional predictive filters discussed previously.

There are several modifications to this basic approach which enhance its operation. First, an array of E_D values is maintained from cycle-to-cycle, and individual elements of this array are computed recursively to reduce the computation required. Second, since the resolution of time shift is about $T/2$ = 0.01667 s, which is not sufficient for accurate high-speed tracking, synthesized subsample data are introduced using linear interpolation between actual measured values. Also, if N_τ is not an integer value, linear interpolation is used to compute the predicted value. With these modifications, simulations show that the correlation prediction technique works quite well for periodic waveforms. As an example, Fig. 11 shows the simulated response of the correlation prediction for a 4.7 Hz sinusoidal input (0.31 times the Nyquist frequency, which is 15 Hz for this example). The target signal is corrupted by noise with a uniform distribution between -0.01 and 0.01. This plot shows that, even at relatively high frequencies, correlation-based prediction can provide accurate and latency-free estimates of target positions. Since the previous motion samples are used as the model, there is no overshoot. Thus, for higher-frequency predictable target motions, such an approach is preferable to the more conventional predictive filters discussed previously. Experimental results using correlation-based prediction for high-frequency periodic motions in conjunction with a polynomial LMSF filter for lower frequency random motions (switching between the two based on the correlation) will be presented in the next section, along with results for the other predictive filters.

3. EXPERIMENTAL RESULTS

In this section, some experimental tracking data are presented to verify the results of the previous section. For these tracking experiments, the redundancy of TRICLOPS (4 degrees of freedom for a 3 degree-of-freedom pointing task) is used to create apparent motion from a stationary target. A fixed target was visually tracked using the tilt and vergence axes while a sinusoidal motion of the TRICLOPS base rotation joint was performed. This results in apparent motion of the target in an arc about the base rotation axis. The frequency of the base motion was increased linearly with time, and the magnitude of the commanded base motion was +/-0.175 rad (+/-10 deg). The target was placed at a distance of 0.9144 m (36 in), so the resulting nearly-sinusoidal motion of the target along the (horizontal) X axis of a coordinate frame fixed with respect to the rotating tilt/vergence platform was 0.175×0.9144 = +/-0.16 m. The response plots will show the actual X position of the target with respect to the tilt/vergence platform and the X component of the tracked position (i.e., where the cameras were pointing) as determined by applying forward kinematics to the tilt and vergence axes. Since image centroid updates are obtained at 30 Hz, $f_{Nyquist}$ = 15 Hz.

First, we will present tracking results for two examples using conventional predictive filters. For these two examples, the frequency of the base motion goes from 0 Hz to 1.5 Hz in 10 s. The example parameters which have been chosen are those which give about the widest range of flat response combined with acceptable behavior in the presence of noise. To meet this

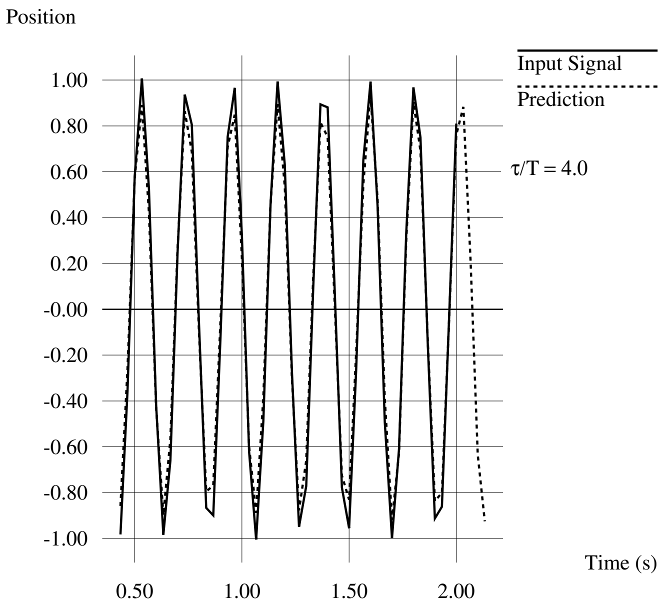

Fig. 11. Simulated response of correlation prediction technique.

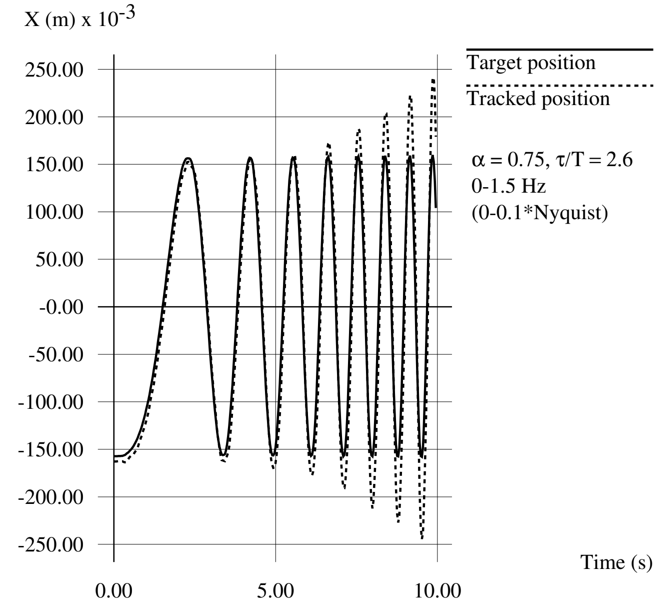

Fig. 12. Experimental performance of α–β–γ filter.

goal, an α–β–γ filter with $\alpha = 0.75$ and a cubic polynomial LMSF filter with $N = 12$ were used. For both of these filters, the amount of prediction applied was $\tau/T = 0.088/0.0333 = 2.6$ samples. This is less than the theoretical amount of prediction required for complete compensation (0.117 s—see section 1.1). Here we have taken advantage of the additional prediction at higher frequencies provided by both of the filters, and have reduced the nominal amount of prediction to reduce the sensitivity to noise.

The measured response of the system with the α–β–γ filter is shown in Fig. 12. Good tracking without overshoot is obtained up to almost 0.75 Hz. The maximum overshoot at 1.5 Hz is about 56% (magnitude ratio of 1.56), which puts it just above curve 4 ($\alpha = 0.75$, $\tau/T = 2.0$) at 0.1 on the frequency scale in Fig. 5. This is expected, since the amount of prediction in the experiment is a bit greater than 2 samples. The response of the cubic polynomial LMSF filter to the same target motion is presented in Fig. 13. For the parameters given, this LMSF filter tracks without overshoot up to about 1 Hz, and the peak overshoot is 40% at 1.5 Hz. This corresponds to a magnitude ratio of 1.4 at 0.1 times the Nyquist frequency, which places it just above curve 3 in Fig. 9 as expected. Both of these filters are clearly having difficulty tracking at 1.5 Hz, and the target is lost from the field of view of the cameras at higher frequencies.

Next, we will present some preliminary results using the correlation-based prediction technique. For the correlation-based prediction, the frequency was increased from 0 Hz to 3 Hz over a period of 60 s. The frequency is increased more slowly in this case so that there will be good correlation over the window of samples. The window size N in (19) was 30, with two synthetic data points between each measured sample. The cubic polynomial LMSF filter described above was used for low-frequency tracking, with the correlation-based tracking becoming active when the minimum E_D dropped below the threshold value (at about 1 Hz). The threshold value for E_D was 0.1 m^2 for this test. Fig. 14 shows a 2-second recording of the tracking performance near the end of this frequency sweep. The motion frequency for the data in the figure is about 2.5 Hz. As shown in the figure, the target motion is tracked very well, without any overshoot, even at this comparatively high frequency (and using the full amount of prediction—0.117 s). There is only a very small amount of latency, due to the fact that the target motion is increasing in frequency, rather than exactly periodic. If the target motion decreases in frequency, there is a corresponding slight overprediction. This effect does not have a significant effect on tracking as long as the frequency changes slowly enough. If the frequency changes quickly, however, E_D will be too large and tracking will be lost.

Although steady-state tracking of periodic targets works quite well with the correlation-based prediction, there are some limitations with the technique in its current state of implementation. The most significant problem is the method used to switch between the LMSF filter and the correlation-based tracking. At frequencies near the transition to correlation-based tracking the system jumps back and forth between the two prediction techniques, which causes some jumpiness in tracking. Also, the experimental system does not track at frequencies as high as simulation studies indicate should be possible.

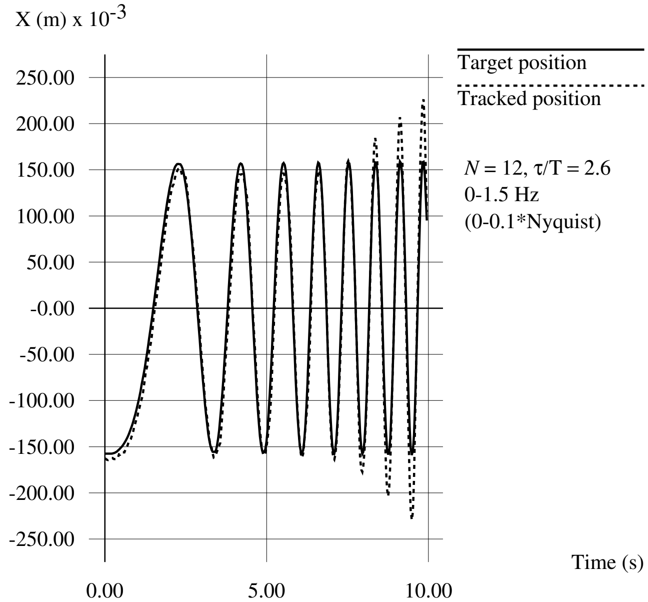

Fig. 13. Experimental performance of cubic LMSF filter.

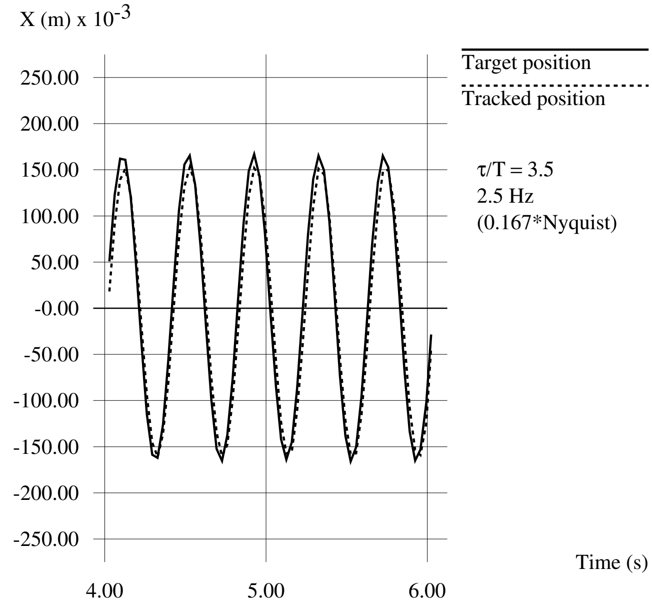

Fig. 14. Experimental performance of correlation prediction.

4. CONCLUSIONS

Prediction is necessary to obtain high-performance tracking in the face of large image acquisition and processing delays. This paper has examined the frequency response characteristics of constant-coefficient Kalman filters and polynomial LMSF filters when used for prediction. It is demonstrated that these filters may be reformulated to be in standard IIR and FIR (for constant-coefficient Kalman filters and polynomial LMSF filters, respectively) forms. All of these predictive filters have a magnitude peak at higher frequencies, which indicates that overshoot will occur. This is true even for the case where the filters are used just for smoothing, without any prediction at all. Larger amounts of prediction and increased smoothing (decreased α for α–β and α–β–γ filters, increased N for polynomial filters) result in lower peaks in the magnitude curve, but also decrease the frequency range for flat magnitude response. Higher-order filters (α–β–γ and cubic polynomial LMSF) are seen to have higher peak magnitude values, but also have a wider bandwidth of flat response. All of the filters except the α–β filter also have a peak in the effective prediction vs. frequency plot. They provide the desired amount of prediction up to about 1-3% of the Nyquist frequency, and then provide a larger amount of prediction until it starts tapering off starting at 6-10% of the Nyquist frequency. The peak in the amount of prediction can serve to extend the effective range of tracking, since the nominal amount of prediction can be reduced (lower frequencies can be tracked with a reduced amount of prediction and still remain within the camera fields-of-view). This results in less overshoot and noise sensitivity, and gives accurate tracking at higher frequencies. If accurate low-frequency tracking is desired, an α–β filter might be more appropriate, since this filter does not have a peak in the prediction vs. frequency curve. The stability and accuracy of the tracking algorithms described in this paper strongly depend on being able to match the correct joint position feedback data to the image data. This is only possible if the image acquisition and processing delay is accurately determined, preferably by direct timing.

A method of tracking higher frequency predictable motions without overshoot or latency based on autocorrelation of the target signal is also discussed. This method is similar to the TSAC approach of Bahill and McDonald[2], except that previous samples of the target, rather than adaptively-determined parameters for a specific waveform type, serve as the model of the motion signal. This method has been used to track sinusoidal target motions in excess of 2.5 Hz. It is desirable to combine the conventional predictive filter with a periodic motion prediction technique, such that both random motions with low frequency content and higher-frequency predictable motions can be tracked successfully with automatic switching between the two techniques. This has been accomplished with some success, but difficulties remain; particularly in performing smooth transitions between the two techniques. For example, if the system is tracking a high-frequency periodic signal which then ceases to become periodic, what should the system do? The prediction provided by a conventional predictive filter cannot be used, since the overshoot will be too great. Probably the best which can be done in this situation is to attempt to follow the periodic trajectory for a few cycles, but then stop tracking altogether. Peripheral vision (provided by the center camera of TRICLOPS) could also play an important role in this case, by keeping the target in the field of view without attempting to track its motions accurately, and by indicating when the target motions have slowed down enough to enable tracking by the vergence cameras. This is an area for future investigation.

5. REFERENCES

1. Allen, P., Yoshimi, B., Timcenko, A., "Real-time visual servoing," In *Proc. DARPA Image Understanding Workshop*, Pittsburgh, PA, 1990.

2. Bahill, A. T., McDonald, J. D., "Adaptive Control Model for Saccadic and Smooth Pursuit Eye Movements," in *Progress in Oculomotor Research*, Fuch, A. F., Becker, W., eds., Elsevier, 1981.

3. Bahill, A. T., "A Simple Adaptive Smith-Predictor for Controlling Time-Delay Systems," *IEEE Control Systems Magazine*, May, 1983, pp 16-22.

4. Bar-Shalom, Y., Fortmann, T. E., *Tracking and Data Association*, Academic Press, Inc., 1988.

5. Blair, W. D., Gray, J. E., Boyd, M. D., "Design Analysis for the Two-stage Alpha, Beta, Gamma Estimator," In IEEE Proc. SOUTHEASTCON '91, Vol. 2, pp. 1050-1054, 1991.

6. Brown, C., "Gaze Control with Interactions and Delays," *IEEE Trans. Sys., Man, & Cybern.*, Vol. 20, No. 1, March, 1990.

7. Brown, C. and Coombs, D. J., "Notes on Control with Delay," University of Rochester Technical Report 387, October, 1991.

8. Brown, C., Coombs, D., and Soong, J., "Real-time Smooth Pursuit Tracking," in *Active Vision*, Blake, A. and Yuille, A., eds., MIT Press, 1992.

9. Fiala, J., Lumia, R., Roberts, K., and Wavering, A., "TRICLOPS: A Tool for Studying Active Vision," to appear in the *International Journal of Computer Vision*.

10. Fiala, J., Wavering, A., "On Visual Pursuit Systems," in publication.

11. Kalata, P. R., "The Tracking Index: A generalized Parameter for α–β and α–β–γ Target Trackers," *IEEE Transactions on Aerospace and Electronics Systems*, Vol. AES-20, No. 2, 1984.

12. Kuo, B. C., *Automatic Control Systems*, Prentice-Hall, Inc., 1982.

13. Kuo, B. C., *Digital Control Systems*, Holt, Rinehart, and Winston, Inc., 1980.

14. Lumia, R., "TRICLOPS: A Tool for Active Vision," *IEEE Robotics and Automation Conference Video Proceedings* (videotape), Atlanta, GA, 1993.

15. McDonald, J. D., Bahill, A. T., "Zero-Latency Tracking of Predictable Targets by Time-delay Systems," *Int. J. Control*, Vol. 38, No. 4, pp. 881-893, 1983.

16. Murray, D. W., Du, F., McLauchlan, P. F., Reid, I. D., Sharkey, P. M., Brady, M., "Design of Stereo Heads," in *Active Vision*, Blake, A. and Yuille, A., eds., MIT Press, 1992.

17. Ng, L. C. and LaTourette, R. A., "Equivalent Bandwidth of a general class of Polynomial Smoothers," *J. Acoust. Soc. Am.*, Vol. 74, No. 3, September, 1983.

18. Parks, T. W., Burrus, C. S., *Digital Filter Design*, John Wiley & Sons, Inc., 1987.

19. Robinson, D. A., "Why Visuomotor Systems Don't Like Negative Feedback and How They Avoid It," in *Vision, Brain, and Cooperative Computation*, Arbib, M. A., Hanson, A. R., eds., MIT, Cambridge, MA, 1988.

20. Smith, O. M., "Closer Control of Loops with Dead Time," *Chemical Engineering Progress*, Vol. 53, No. 5, pp. 217-219, 1957.

21. Wavering, A. J., Fiala, J. C., Roberts, K. J., Lumia, R., "TRICLOPS: A High-performance Trinocular Active Vision System," In *Proc. IEEE Conference on Robotics and Automation*, Vol. 3, pp. 310-317, Atlanta, GA, 1993.