A comparison of foveated acquisition and tracking performance relative to uniform resolution approaches

Shaun Dubuque[a], Thayne Coffman[a], Paul McCarley[b], A.C. Bovik[c], C. William Thomas[a]

[a]21st Century Technologies, 4515 Seton Center Parkway Ste. 320, Austin, Texas 78759-5731, USA;
[b]Air Force Research Laboratory, Munitions Directorate (AFRL/RWGI);
[c]Center for Perceptual Systems, Dept. of Electrical and Computer Engineering,
The University of Texas at Austin, Austin, TX 78712-1084, USA;

ABSTRACT

Foveated imaging has been explored for compression and tele-presence, but gaps exist in the study of foveated imaging applied to acquisition and tracking systems. Results are presented from two sets of experiments comparing simple foveated and uniform resolution targeting (acquisition and tracking) algorithms. The first experiments measure acquisition performance when locating Gabor wavelet targets in noise, with fovea placement driven by a mutual information measure. The foveated approach is shown to have lower detection delay than a notional uniform resolution approach when using video that consumes equivalent bandwidth. The second experiments compare the accuracy of target position estimates from foveated and uniform resolution tracking algorithms. A technique is developed to select foveation parameters that minimize error in Kalman filter state estimates. Foveated tracking is shown to consistently outperform uniform resolution tracking on an abstract multiple target task when using video that consumes equivalent bandwidth. Performance is also compared to uniform resolution processing without bandwidth limitations. In both experiments, superior performance is achieved at a given bandwidth by foveated processing because limited resources are allocated intelligently to maximize operational performance. These findings indicate the potential for operational performance improvements over uniform resolution systems in both acquisition and tracking tasks.

Keywords: Foveated imagery; biomimetic image processing; multi-resolution image processing; automatic target acquisition; automatic target tracking; visual search; variable acuity

1. INTRODUCTION

Foveated imagery is imagery with spatially varying resolution or acuity. In contrast to almost all imagery in use today, some areas of foveated images (the foveae) are processed at high resolution, while other areas (the periphery) are processed at lower resolution. The most advanced natural visual systems, including those of humans, are all built on active foveated imagery [1]. By using spatially-varying resolution, foveated imagery can simultaneously monitor wide fields of view, maintain high resolution on areas of interest, and minimize processing and bandwidth requirements. These same traits offer the potential for superior performance in artificial vision systems. Automatic target acquisition, tracking, and recognition (collectively called targeting) are three key technical applications for these systems.

New digitally foveating cameras offer the potential to directly generate and exploit variable resolution imagery and thereby improve the performance of modern vision systems [2][3]. These cameras generate foveated imagery in visible light or infrared modalities directly on the focal plane array, typically by sharing charges between adjacent pixels during image formation. This reduces the bandwidth required to transfer the image off the focal plane, allowing cameras with frame rates approaching 4000 Hz.

However, a set of technical barriers still prevents the development and use of high-speed foveated targeting systems. Theory has been developed to guide the use of foveated imagery in applications for human consumption (e.g., compression or teleconferencing), but there is a relative lack of theory to guide the development of foveated targeting systems. Further, state-of-the-art techniques for foveated image formation, resolution control, acquisition, tracking, and recognition are all too simple and too brittle to be effective in many scenarios. Acquisition algorithms are not robust, and almost exclusively use motion as a cueing indicator, while ignoring target appearance, scene content, and context. Even the underlying tracking theory, including current formulations of the ubiquitous Kalman filter, is demonstrably insufficient to exploit the capabilities of variable resolution and variable frame rate vision systems.

While more and more in the research community are exploring foveated vision [4][5][6], the majority of commercial and military targeting systems still rely on uniform resolution imagery. Compelling evidence must be presented that foveated imagery can achieve better targeting performance in order for it to be adopted outside the laboratory. This is the driving goal behind the research and experiments described within. These experiments begin to fill voids in the methodologies for analyzing foveated targeting systems and for choosing between foveated and uniform resolution imagery in targeting systems.

2. FOVEATED ACQUISITION EXPERIMENTS

2.1 Methodology

Our acquisition experiments tested the performance of localizing a hidden but known target. A single target was placed randomly at an unknown location in a noisy background. Background images were 256x256 pixels, and were drawn from a parameterized family of random processes that encompass white, pink, and red noise. As white and red noise is more common within image models we chose to keep our experiments focused on these. For review, white noise has a flat frequency spectrum ($S(f) \propto 1/f^0$), pink noise has a spectrum inversely proportional to frequency ($S(f) \propto 1/f^1$), and red noise has a spectrum inversely proportional to frequency squared ($S(f) \propto 1/f^2$). The experiments embedded targets in a variety of background models to study their relative effects. Tested models included static white noise, time-varying white noise, static red noise, and static red noise plus time-varying white noise.

The targets used were comprised of a single two dimensional Gabor wavelet function. The real-valued cosine formulation of a Gabor wavelet is a Gaussian envelope modulated by a 2D sinusoid of particular frequency and orientation [7]. The targets were thus defined by

$$W(x_0, y_0, \lambda_x, \lambda_y, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y}\left(\exp\left(-\pi\left[\left(\frac{x-x_0}{\sigma_x}\right)^2 + \left(\frac{y-y_0}{\sigma_y}\right)^2\right]\right)\right)\cos(\lambda_x x + \lambda_y y),$$

where (x_0, y_0) define the spatial location of the target center, (λ_x, λ_y) define the (single) 2D target frequency, and (σ_x, σ_y) define the width of the Gaussian envelope. The targets' spatial extents were limited to a 40x40 area around its target center, and Gaussian envelope width was fixed at $(\sigma_x, \sigma_y) = (13.5, 13.5)$. Target frequency (λ_x, λ_y) was varied.

We believe that results and trends observed on these fundamental targets can be extended to more complex targets and plan to explore these within future investigations. Gabor wavelets as fundamental targets is motivated by research which indicates that early stages of the human visual system respond almost directly to Gabor wavelet input patterns, and that some of the specialized aggregator neurons in the visual pathway are tuned to specific Gabor locations and frequencies [8]. Given this, Gabor wavelets can be seen as fundamental basis vectors or features into which the early human visual system decomposes its input and perceives the images presented to it. **Fig. 1** shows examples of Gabor wavelet targets embedded in two different background noise models.

In our acquisition experiments we examine the performance of both full resolution and foveated resolution approaches. In order to emulate the output of a foveated vision system we filter the acquisition images using a bank of low pass filters. Specifically, we pass the full resolution image through a bank of 30 low pass filters with cutoff frequencies linearly spaced on the range [0, 1]. Each low pass filter is designed by applying a Hamming window to an ideal low pass filter with the desired cutoff frequency [9]. As part of its operation, the foveated acquisition algorithm specifies the position (x_f, y_f) of the fovea. Together with a set of foveation parameters, that position defines the desired bandwidth (or equivalently, the amount of filtering) at each pixel. The output of the filter with cutoff frequency closest to the desired bandwidth (which lies on the range [0, 1]) at each pixel then defines the value of the foveated image at that pixel. The result is an image approximating the result of a vision system with resolution that varies smoothly over space. We are aware of other foveation methods, such as those based on halftoning and pixel sharing [10], but chose not to explore them within this study.
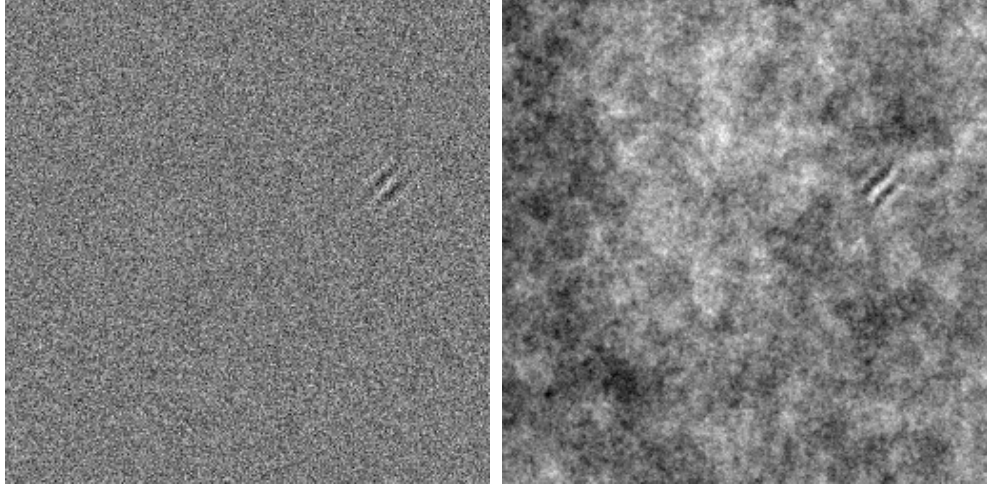
Fig. 1. Full resolution images of Gabor wavelets embedded in white (left) and red (right) noise.
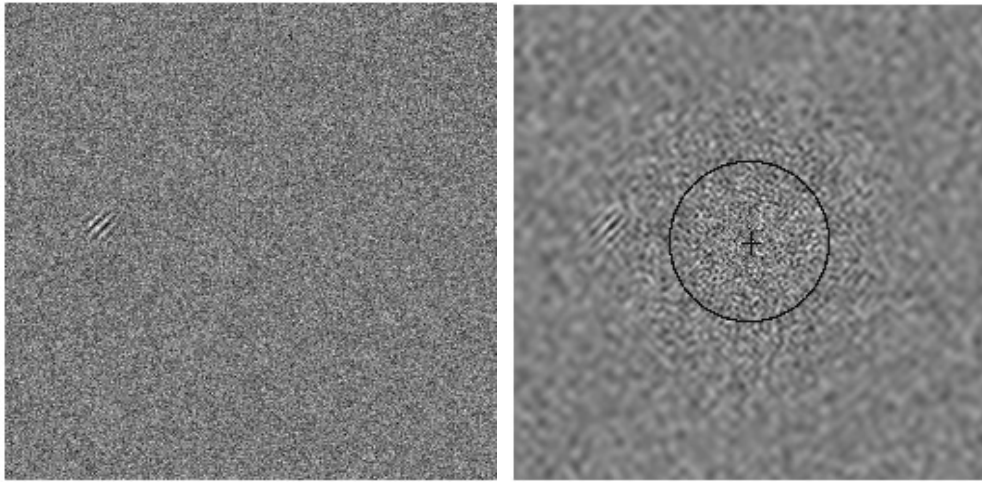

Fig. 2. (Left) Full resolution image of Gabor target in white noise. (Right) Foveated image.

We define the desired bandwidth as follows:

For a fovea placement at (x_0, y_0) with fovea width σ, falloff exponent α, and a minimum desired bandwidth value ε, the desired bandwidth at each pixel (x, y) in the foveated image is given by

$$\rho(x, y) = \sqrt{(x - x_f)^2 + (y - y_f)^2} \,,$$

$$BW(x, y) = BW(\rho(x, y)) = \max\left\{\exp\left(-\left(\frac{\rho}{\sigma}\right)^\alpha\right), \varepsilon\right\}.$$

Experiments used a falloff exponent of $\alpha = 1.0$, which matches empirical studies of the human visual system [1]. Additionally, minimum desired bandwidth was set to $\varepsilon = 0.2$. **Fig. 2** shows an example of full-resolution and foveated images of a Gabor wavelet target in white noise. The half power bandwidth (HPBW) is defined as the radius of the circular region near the fovea for which desired bandwidth is at least 0.5. In the foveated image, the fovea center is marked with a plus and the HPBW edge is marked with a circle.

The foveated approach uses mutual information between the known target and foveated observed images to select fovea placements. Mutual information measures the amount of statistical dependence of two random variables *X* and *Y*. Thus, the more dependent the two variables, the more redundant the information they provide and therefore the greater their mutual information, *I(X;Y)*. Mutual information is defined in terms of the entropy *H(X)* and *H(Y)* in each random variable and the conditional entropy of one variable given the other, *H(X|Y)*. It is defined as

$$I(X;Y) = H(X) - H(X \mid Y)$$

or equivalently, for discrete random variables, as

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right),$$

where $p(x,y)$ is the joint probability mass function of the two variables, and $p(x)$ and $p(y)$ are the individual probability mass functions [11].

Our foveated acquisition approach centers the fovea on the pixel with largest mutual information between the target and the observation. However, because the approach is designed to process foveated *video* inputs from an active vision system, acquisition takes place over multiple consecutive frames and fovea placements. New fovea placements are not allowed to fall within the HPBW region of any previous placement to prevent revisiting previously explored regions. This restriction is valid under the assumption that the target does not move, appear, or disappear between foveated video frames as is the case within our experiments.

In addition to fovea placement via mutual information, we evaluated a trivial random fovea placement algorithm which places fovea at successive pixels with equal probability, subject to the same heuristic for avoiding previous placements. This approach serves as a performance baseline with which we can compare other algorithms' performance. Fovea placement approaches can be said to add value only when it can be shown that they outperform random placement (even random placement is effective to some degree in many cases).

Performance was evaluated using detection probability metrics and counts of the number of fovea placements required before acquisition was achieved. The primary metric used for evaluation was percent acquisition, which measured the probability of placing a fovea near the location of the target within a limited number of attempts. Our experiments were based on the simplifying assumption that once a fovea was placed close enough to the target – specifically, such that the target falls within the HPBW region of the fovea – a notional target recognition algorithm may be applied which detects the target with 100% probability. This very significant assumption means that our experiments solely test the fovea placement and target localization performance of the approaches. The assumption was deemed acceptable due to the fact that once a high-resolution fovea is placed on the target, existing full resolution acquisition and recognition approaches can be applied making comparisons between the two negligible. As discussed, the percent acquisition metric is the percentage of tests in which the fovea was successfully placed near the target in fewer than 10 fovea placements. No consideration is given in this metric to whether the target was acquired on the first, second, or tenth placement. As noted, a second metric was used which computed the average number of fovea placements required to place the fovea near the target. This metric was only computed in cases where the target was successfully acquired.

2.2 Experimental Results

The first set of experiments evaluated foveated acquisition performance on targets embedded in static noise with varying signal-to-noise (SNR) ratios. The purpose of this experiment was to discover possible trends in the effects of SNR on acquisition performance. In this experiment, SNR was varied from -24 dB to -2 dB, the fovea half-power bandwidth (HPBW) radius was held constant at 43 pixels, and target frequencies were varied through approximately 29, 58, 86, and 115 cycles/image. For the results presented below, each data point represents 100 experiment executions with identical parameters and randomized target locations.

Results are shown in **Fig. 3**. As was anticipated, foveated acquisition using mutual information performs better as SNR increases. Also as expected, random fovea placement is independent of SNR. For the lowest frequency targets, any SNR above -20 dB results in nearly perfect acquisition results. However, for the highest frequency targets, even strengths of -2 dB do not provide reliable localization. The minimum SNR for detection appears to be monotonic in target frequency.
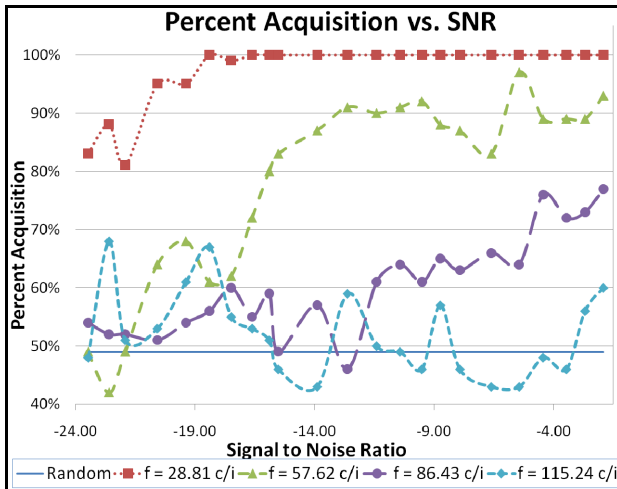
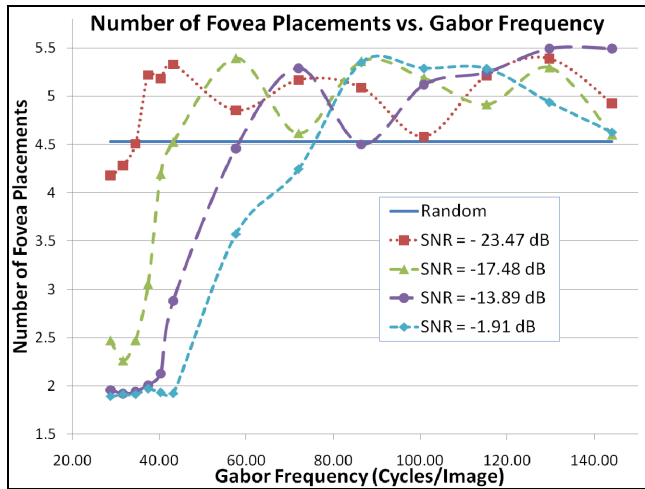Fig. 3. Acquisition performance vs. SNR for various target frequencies in static white noise background.

Fig. 4. Number of fovea placements vs. target frequency and SNR in static white noise background.

Our next set of experiments further explored the trend relating acquisition performance to target frequency. Within this experimental setup target frequencies were varied from approximately 30 to 140 cycles/image, SNR was varied through -24, -17, -14, and -2 dB, and fovea HPBW was held constant at 43 pixels.

The average number of fovea placements required for successful acquisition is shown in **Fig. 4**. For easy acquisition tasks (those with high SNR and low target frequency), successful acquisition required far fewer placements than the random placement strategy. Ignoring the first placement, which is always in the image center, a single placement based on mutual information consistently localized the target. With more difficult acquisition tasks the placement strategy requires *more* placements than the random strategy on average, even when it continues to provide better acquisition performance. Eventually, as the strategy becomes ineffective and provides no value over random placement, its number of placements becomes similar to that of the random strategy. We believe that the average number of fovea placements for successful acquisition may provide a useful indication of the relative difficulty of the acquisition task.

The trends detected were neither strictly monotonic nor entirely consistent. In some cases there was enough variation in the results to question the existence of those trends. However, future work including a similar analysis but with significantly more repetitions for each parameter combination should reduce the impact of random perturbations on the underlying trends.

Our final set of experiments compared the performance of the foveated acquisition approach relative to two notional uniform resolution approaches. In order to evaluate full resolution and foveated acquisition systems on equal footing, the systems were constrained to consume equal bandwidth, and resulting performance was measured relative to time in seconds instead of number of frames. Foveated cameras and processing can often achieve significantly higher frame rates than full resolution systems due to the reduced data volume per frame, so measuring performance relative to frame index can incorrectly overstate the relative performance of full resolution systems.

The equal-bandwidth constraint was based on the simplifying assumption that a sensor's frame generation time is linear in the number of pixels per image. Given a frame size in pixels and a frame rate we calculate bandwidth as the number of pixels per frame times the number of frames per second. We can use the desired bandwidth function $BW(x, y)$ to approximate the frame rate of a foveated camera that would consume identical bandwidth. A foveated image modeled by low pass filtering full resolution images can be critically sampled at each spatial location at a density proportional to $BW(x, y)$. Thus, we can define a measure of relative bandwidth as the sum of $BW(x, y)$ over the foveated image divided by the number of pixels in the full resolution image. For full resolution images we have $BW(x, y) = 1$ at all locations.

As an example, assume full resolution images are generated at 30 frames per second (30 Hz), HPBW is set to 50 pixels, and minimum desired bandwidth is $\varepsilon = 0.2$. In the worst case of placing the fovea directly in the center of the image in every frame, each foveated frame would require 30.93% of the bandwidth of a full-resolution frame. Thus, a foveated

camera generating frames at 97 Hz would consume the same bandwidth as a full-resolution camera generating frames at 30 Hz.

Again, targets were modeled as a Gabor wavelet. For these experiments, a target frequency of 58 cycles/image was used and embedded within a static white noise background with SNR of -14 dB. This SNR was selected by inspection to be just over the threshold for human detection. Using a fovea HPBW of 50 pixels, our previous results indicated that the foveated approach successfully localized the target in 96% of the standard 10-placement acquisition experiments. From the details of the individual 100 runs, we reconstructed the marginal target acquisition rate at (or before) each of the successive 10 foveated frames. Doing so provided us with an estimate of the acquisition probability for any number of placements (up to 10) as well as the estimated foveated system frame rate in Hz. Using these together we computed an estimate of foveated acquisition performance as a function of time instead of frame index.

Interpreting the acquisition experiment data in terms of time produces **Fig. 5**, which shows the estimated foveated target acquisition performance vs. time in seconds, relative to two uniform resolution systems that consume identical bandwidth. Full resolution acquisition performance was modeled to have an independent and constant probability of successful acquisition per frame (which is extremely generous to the full resolution system). As shown in **Fig. 5**, if the full resolution system had a per-frame detection rate near 55%, then at almost any given time in seconds the foveated system exhibited a higher likelihood of target acquisition than the full resolution system. If the full resolution system instead had a per-frame detection near 90%, it would exhibit a higher likelihood of target acquisition at any time past 33ms. Even in this case, however, the foveated system would still exhibit roughly a 50% chance of acquiring the target before the full-resolution system. This analysis allows us to identify the conditions under which a particular foveated system can (or cannot) be expected to acquire targets faster than an equal-bandwidth full resolution system based on expected algorithm performance.

In conclusion, the foveated target acquisition experiments revealed a number of interesting insights with significant implications for the deployment of future foveated target acquisition systems. Under a range of assumed full resolution acquisition system performance, foveated systems can be expected to outperform full resolution systems of equal bandwidth. In some cases the foveated system always exhibits a higher likelihood of acquisition by time in seconds, but in cases where the full resolution system eventually exceeds the probability of detection of the foveated system, the foveated system can still have a significant likelihood of acquiring the target before the full resolution system. Performing analysis based on equal-bandwidth assumptions and performance comparison relative to time in seconds (as opposed to frame index) we can identify the conditions under which foveated systems can or cannot be expected to outperform equal-bandwidth full resolution systems.
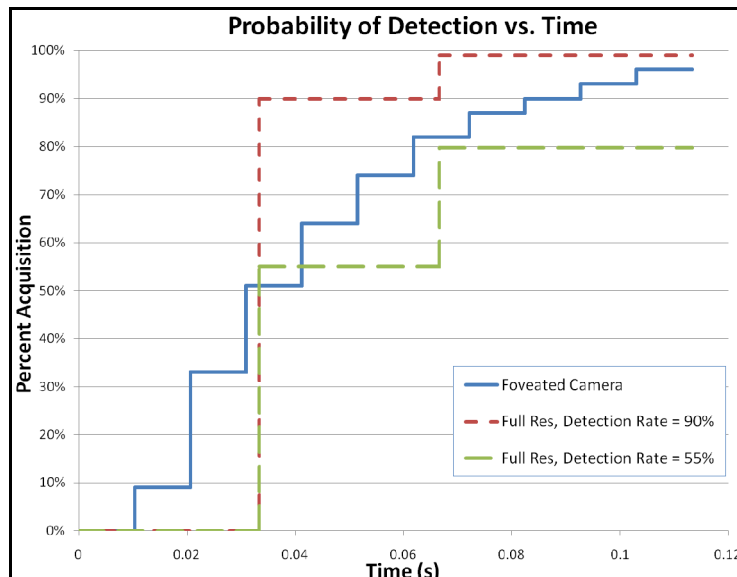

Fig. 5. Estimated foveated and full-resolution acquisition performance as a function of time.

3. FOVEATED TRACKING EXPERIMENTS

The purpose of our tracking experiments was to uncover trends shaping the performance of foveated tracking systems and to compare their performance against full resolution systems.

3.1 Methodology

Our tracking experiments simulated visual tracking of multiple targets using Kalman filters. We evaluated foveated and uniform resolution imaging approaches through appropriate observation models which described localization accuracies for each. In these experiments we use thee different observation models. Additionally, we assume the existence of a target acquisition algorithm which generates noisy observations of the target positions in each frame. Foveated camera models assume the existence of a *fovea scheduling* algorithm which selects, for each frame, a single target to be scheduled for fovea placement. Aside from model specific parameters, all of the camera models were designed to share the same intrinsic and extrinsic parameters including a field of view of 0.5 radians and a maximum resolution of 2000x2000 pixels.

The *uniform camera model* provides uniform resolution observations across the field of view. For Kalman filtering, this results in constant observation error covariance for all locations within an image. Uniform resolution cameras are nominally assumed to operate at 30 Hz, without loss of generality.

The *simplified foveated camera model* emulates variable resolution imagers with two discrete resolutions: high resolution foveal resolution and low resolution peripheral resolution. Observations of a target are defined to either occur with the fovea centered directly on the target (*foveal observations*) or with the fovea far from the target (*peripheral observations*). The result of this is that the selected target is observed at maximum (foveal) resolution and all other targets are observed at minimum (peripheral) resolution. This would be equivalent to placing an infinitely small fovea (HPBW ≈ 0) over a single target in each frame. Peripheral resolution is defined in terms of a downsampling factor (DF). The periphery is assumed to be spatially downsampled by the factor DF relative to the foveal resolution. Observation error covariance is directly proportional to DF, as discussed below.

The *extended foveated camera model* emulates the smoothly varying resolution foveation applied in our acquisition experiments. Foveas within this model are of non-zero size and exhibit gradual falloff from maximum to minimum resolution. Fovea placement may not precisely center the fovea over the scheduled target. Targets nearby the scheduled target will be observed at higher resolution and thus exhibit lower observation error covariance.

In our experiments, the observation models described above were used to observe simulated targets. Our targets were defined with constant velocity (linear motion) and were modified at constant intervals by the application of Gaussian noise to both target position and velocity. The noise magnitude added to target position was chosen to be a factor of 10 smaller than the noise added to velocity. This was chosen through experimentation to produce interesting and non-trivial target motion. **Fig. 6** shows example target paths for a 5-second period in one experimental run.
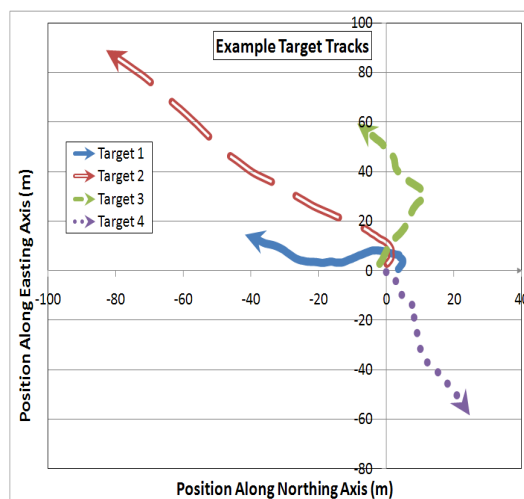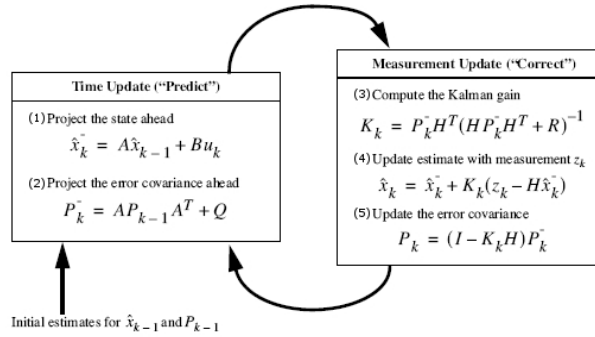

Fig. 6. Four example target paths.

Fig. 7. Kalman filter flowchart and principle equations [12].

A Kalman filter [12] was used within our tracking experiments to auto-regressively estimate target position and velocity based on observations generated by our observation models. Target process models were derived from the constant velocity additive Gaussian approach discussed above. In order fuse foveal and peripheral observations into a consistent state update model we had to define a simple set of Kalman filter extensions. These extensions specifically relate to Equation (3) in the Kalman "Update" phase illustrated in **Fig. 7**. This equation guides the computation of the Kalman filter gain. This formulation of gain weights the influence of the previous state and new observations according to their associated covariances and computes the gain that minimizes the mean square error of the resulting state estimate. For the uniform camera model this observation variance is constant across the field of view. For the foveated camera models, observation variance is assumed to change according to the resolution at the position of the observation.

When tracking with a uniform camera model, the observation covariance is defined as

$$\mathbf{R} = \sigma_m{}^2 \cdot \mathbf{I} ,$$

where **R** is the observation covariance matrix defined in Equation (3) of **Fig. 7**, $\sigma_m{}^2$ is the observation variance in meters^2, and **I** is the 2x2 identity matrix. When using a foveated camera model (either simple or extended), the observation covariance matrix can be defined as

$$\mathbf{R} = \sigma_{m,k,r,c}{}^2 \cdot \mathbf{I} ,$$

where **R** is the observation covariance matrix defined in Equation (3) of **Fig. 7**, $\sigma_{m,k,r,c}{}^2$ is the observation variance in meters^2 for observation number k at image location (r,c), and **I** is the 2x2 identity matrix. In this sense we constantly recompute the observation covariance as a function of instantaneous resolution for each separate observation. For the simple foveated camera model, each observation is described with one of two observation covariance matrices, $\mathbf{R_F}$ and $\mathbf{R_P}$, which correspond to observations from the high resolution foveal regions ($\mathbf{R_F}$) and the low resolution peripheral regions ($\mathbf{R_P}$), respectively. Because the simple foveated camera model only supports these two resolutions we can precompute these matrices and select the appropriate observation covariance when computing the Kalman gain for each measurement update.

The extended foveated camera model lacks the discrete resolution choices of the simple model. As a result of this we must compute observation covariance matrices for each measurement at each time step and at each position. The resolution of a measurement taken with the extended foveated camera model is governed by the distance of the observation location from the fovea center, the half power bandwidth, and the minimum bandwidth ratio. This resolution in turn guides the amount of observation variance and subsequently the observation covariance **R.** Ongoing work continues to develop our extended foveated camera model and analyze its performance. All results presented here were generated using the simplified foveated camera model.

In our experiments we used a "round robin" fovea scheduling strategy. For experiments with one target the fovea was centered directly on the target in each frame. For experiments with N>1 targets, the fovea was placed on each target in turn for one frame resulting in cycles of one foveal observation followed by (N-1) peripheral observations.

For these experiments performance was measured as the mean Euclidean distance between the estimated target positions and the ground truth positions at each time step.

3.2 Experimental Results

Our first set of tracking experiments were designed to explore performance when tracking multiple targets with a single-fovea operating over a variety of frame rates. Fovea scheduling effects are important to observe as the time between foveal observations is proportional to the number of targets being tracked. Although it must schedule fovea placements, the foveated system still benefits from a higher frame rate. Peripheral observations of each target collected between foveal observations can be shown to more than make up for the lower frequency of foveal observations.

Fig. 8 shows the mean absolute estimate error in meters for a single-fovea system tracking varying number of targets at varying frame rates. In the figure legend, "Uniform" refers to a uniform resolution system and "Fov" refers to a foveated system. The number of targets and the assumed base observation variance (named "ObsVar" in the figure and σ_m^2 in previous discussion) are also given. In this experiment we used a simplified camera model with a downsampling factor DF=4.

As can be seen in **Fig. 8**, the tracking performance of the uniform resolution system depends only on frame rate and ObsVar, and is unaffected by the number of targets in these experiments. This can explained simply by the fact that all observations are made at maximum resolution regardless of target number. As expected, the performance of a foveated camera operating at a given frame rate decreases with increasing number of targets. For any fixed frame rate, foveated system performance on multiple targets is bounded below by the performance of a uniform resolution system operating with identical observation variance. At the same frame rate, foveated system performance can be shown to be bounded above by the performance of another uniform resolution system operating with observation variance equal to

$$\sigma_{FR}^2 = DF \cdot \sigma_{Fov}^2 ,$$

where σ_{Fov}^2 is the base (foveal) observation variance in the foveated system, DF is downsampling factor, and σ_{FR}^2 is the full resolution system observation variance. These results match intuition as a full resolution system operating at σ_{FR}^2 yields equivalent observations to a foveated system that makes only low resolution peripheral observations.

Our second set of experiments explored the effects on foveated system performance of changing the base observation variance. Results from this experiment were consistent with expectations, and are not shown here. The performance of the Kalman tracker is directly related to the base observation variance, and any increase in observation variance directly increases the tracking error. These results yielded performance curves which can be used to predict the performance of foveated systems under a variety of parameters.
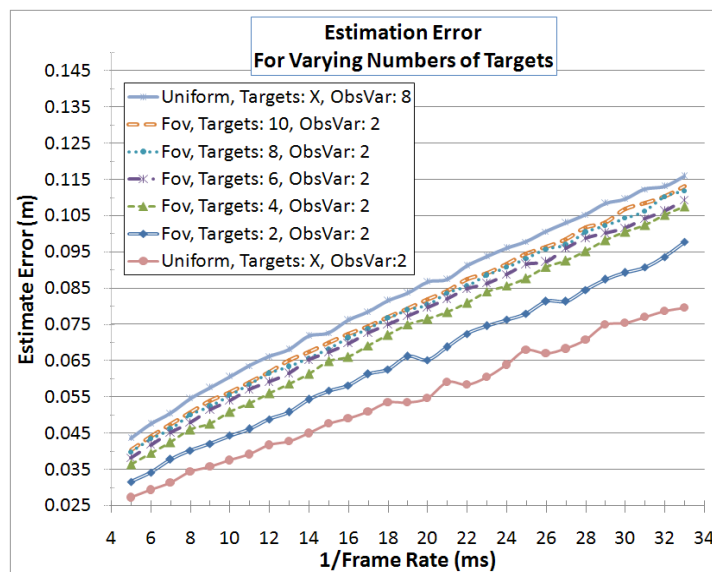

Fig. 8. Estimation error vs. frame rate for varying numbers of targets.

As with acquisition experiments, the comparison between full resolution and foveated systems is most accurate when considering the bandwidth consumed by the various systems. This style of comparison provides insight into how best to maximize system performance when bandwidth is scarce, or when frame rate is tied to bandwidth, as is the case with many modern foveating cameras and any algorithms that scale linearly. As before, bandwidth is defined as the number of pixels per second for both the uniform resolution and simple foveated camera models. Also as before, a full resolution system's bandwidth is defined as resolution (in pixels) times frames per second. Under the simple foveated camera model, the infinitely small fovea and downsampled peripheral resolution yields a number of pixels per frame that is related to the full resolution camera model by

$$N_{Fov} = \frac{N_{FR}}{(DF)^2}.$$

For a foveated system operating at $DF=2$, the required bandwidth is roughly one fourth that of a uniform camera operating at the same frame rate. Conversely, a foveated system operating at $DF=2$ could theoretically achieve a frame rate four times higher than that of a full resolution camera while consuming the same bandwidth. While maximum framerates would be slightly lower in practice due to additional processing requirements for foveation and non zero fovea sizes we submit that similar results would be achieved.

For this final experiment we compared the performance of both a full resolution and simple foveated system while keeping the target number constant at 4. The results from this experiment are shown in **Fig. 9** below. For each system, performance was measured with base observation variance set to 0.5, 1, 2, 4, 8, and 16 pixels^2. This observation variance is constant for all observations in the full resolution system and in-fovea observations within the foveated system. Observation variance for peripheral observations within the foveated system is related to base observation variance through the downsampling factor as discussed earlier. Frame rate (in milliseconds) and downsampling factor DF are specified within the figure for each system simulated.

The uniform resolution system was assumed to operate at 30 Hz (33 ms/frame). As a result, it generates 256·256·30=1,966,080 pixels/second, which we use as our standard for comparing relative bandwidth. Three different parameterizations of the simple foveated camera model were evaluated. The first operated at $DF=2$ and 62.5 Hz (16 ms/frame). This was a lower frame rate than the 120 Hz that an equivalent-bandwidth foveated system could theoretically achieve. As a result, this foveated system requires only 52% of the bandwidth of the uniform resolution system, but as seen in **Fig. 9**, it still outperforms the uniform resolution system. The second parameterization had $DF=3$ and 143 Hz (7 ms/frame), and consumed only 53% of the uniform resolution bandwidth while again outperforming it. The final parameterization used DF=4, operated at 250 Hz (5 ms/frame), and performed better still while again consuming only 52% of the original bandwidth.
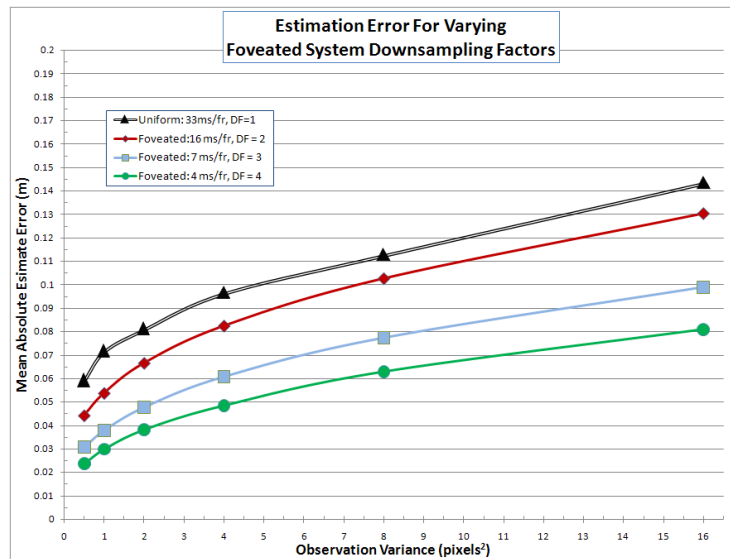

Fig. 9. Estimation error for varying foveated system downsampling factors.

We can observe through these experimental results that under reasonable abstractions, foveated systems can often outperform full resolution systems in multiple target tracking tasks due to their ability to make target observations at a higher frame rate. As shown in **Fig. 9**, even when considerably reducing the frame rate (and bandwidth consumption) expected from a foveated system based on its downsampling factor, the foveated tracking approach generates more accurate target tracks than the full resolution approach. The more extreme the downsampling the faster the achievable frame rate and the lower the tracking error. All three foveated systems outperform the uniform resolution system, independent of observation variance or downsampling factor. Based on this collection of results, we argue that foveated tracking systems can be expected to *always* outperform full resolution systems that consume equivalent bandwidth. This is caused by the significant increase in the frame rate achievable by a foveated system, which enables a more efficient allocation of resolution and bandwidth.

These results were further reinforced by another set of simulations (results not shown) performed under relatively conservative assumptions. In these, foveated systems outperformed uniform resolution systems of equivalent bandwidth while tracking up to 10 targets. For example, a foveated system with *DF=2* tracking two targets at 40 Hz produces more accurate tracks than a full resolution system operating at 30 Hz. A foveated system with *DF=2* tracking 10 targets at only 55 Hz produces more accurate tracks than a full resolution system operating at 30 Hz. These increases in frame rate are well within the capabilities of today's foveated systems (and processing algorithms) because of the substantial bandwidth reduction that foveation provides. Further, the foveated systems require only 47% and 65% of the uniform resolution system's bandwidth, respectively, while still providing superior track accuracy.

4. DISCUSSION

This work provides important early steps towards improved exploitation and extended use of foveated imagery in visual target acquisition and tracking systems. A set of simulations uncovered a variety of positive indictors for the performance improvements to be gained by developing foveated acquisition and tracking systems.

Acquisition experiments demonstrated the operation of a simple foveated acquisition approach based on mutual information between a known but hidden target and the input image. For targets with at least one dominant low frequency, significant foveation can be applied to the input image while maintaining acceptable acquisition performance. Higher frequency targets (or equivalently, smaller targets or sensors with lower maximum acuity) and targets in stronger background noise require the use of less foveation and thus more bandwidth. Targets were marginally more difficult to acquire in red noise than in white noise. A key finding was that under a set of simplifying assumptions, the measured foveated acquisition system results would outperform a variety of notional full resolution systems that consume equivalent bandwidth. If system performance is limited by communication bandwidth or computational complexity, foveated acquisition systems may offer improved performance over full resolution systems.

Tracking experiments demonstrated the operation of a simple variation of the Kalman filter that optimally fuses observations from a foveated camera. Matching intuition, the error in estimated positions for a foveated system increases with increasing observation variance, increases with increasing number of targets, increases with increasing peripheral downsampling, and decreases with increasing frame rate. Under our models, the increasing frame rate, however, dominates the effects of increased downsampling, resulting in two key findings. First, when tracking multiple targets, the expected performance of a foveated system can be bounded above and below by two uniform resolution systems operating at the same frame rate. Second, and most critically, a foveated multi-target tracking system can be expected to always outperform a full resolution system that consumes equivalent bandwidth and has analogous observation variances. Even though the foveated system's peripheral observations are less accurate than the full resolution system, the bandwidth reduction and resulting frame rate increase caused by foveation are the dominating factors and result in superior performance.

These results clearly indicate that foveated systems offer promise for both target acquisition and multi-target tracking tasks, and could be applied to either military or commercial applications. However, this work only begins to model and quantify the benefits that foveated targeting systems can offer. Future work is required, as the results presented here are from initial studies in an ongoing line of research. Future work should include the evaluation of more complex and realistic target, background, and sensor models (particularly the extended foveated camera model), as well as the evaluation of alternative acquisition, tracking, and foveation approaches. Further performance benefits should also be achievable through the development of adaptive foveated exploitation techniques, which will tailor the fovea shape or search strategy to the particular targets and backgrounds being observed.

REFERENCES

[1]   Field, D.J., "Relations Between the Statistics of Natural Images and the Response Profiles of Cortical Cells," *Journal of the Optical Society of America A* 4, 2379-2394 (1987).

[2]   McCarley, P.L., Massie, M.A., and Curzan, J.P., "Large format variable spatial acuity superpixel imaging: visible and infrared systems applications," *Proc. SPIE 5406*, 361-369 (2002).

[3]   McCarley, P.L., Massie, M.A., and Curzan, J.P., "Foveating infrared imaging sensors," *Proc. SPIE 6660A,* 1-14 (2007).

[4]   Tavassoli, A., "Discovery and Representation of Human Strategies for Visual Search," University of Texas at Austin, Department of Electrical and Computer Engineering, (2007).

[5]   Wang, Z., Lu, L., and Bovik, A. C., "Foveation scalable video coding with automatic fixation selection," *IEEE Transactions on Image Processing* 12, (2003).

[6]   Klarquist, W.N., and Bovik, A.C., "FOVEA: A foveated, multi-fixation, vergent active stereo system for dynamic three-dimensional scene recovery," *IEEE Transactions on Robotics and Automation* 14, 755-770 (1998).

[7]   Bovik, A.C., ed., [The Essential Guide to Image Processing], Elsevier Science & Technology, (2009).

[8]   Pollen, D., and Ronner, S., "Visual cortical neurons as localized spatial frequency filters," *IEEE Transactions on Systems, Man, and Cybernetics* SMC-13, 907-916 (1983).

[9]   Oppenheim, A.V., Schafer, R.W., and Buck, J.R., [Discrete-Time Signal Processing (2nd Edition)], Prentice Hall, (1998).

[10]   Coffman, T., Evans, B.L., and Bovik, A.C., "Halftoning-Inspired Methods for Foveation in Variable-Acuity Superpixel Imager Cameras," *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, (2005).

[11]   Guiasu, S., [Information Theory with Applications], McGraw-Hill, (1977).

[12]   Welch, G., and Bishop, G., "An Introduction to the Kalman Filter," University of North Carolina, Department of Computer Science, TR95-041, (1995).