Anti-Temporal-Aliasing Constraints for Image-Based Feature Tracking Applications With and Without Inertial Aiding

Michael J. Veth, *Senior Member, IEEE*, Richard K. Martin, *Member, IEEE*, and Meir Pachter, *Fellow, IEEE*

***Abstract*—Image-aided navigation techniques can determine the navigation solution (position, velocity, and attitude) by observing a sequence of images from an optical sensor over time. This operation is based on tracking the location of stationary objects in multiple images, which requires solving the correspondence problem. This paper enhances previous research efforts to characterize the correspondence problem using fundamental optical principles and statistical temporal sampling theory by including a rigorous derivation of the Nyquist constraint in pixel space. This development results in a general temporal sampling constraint and reveals the essential connection between the deleterious effects of temporal aliasing and the ambiguities that plague the correspondence search problem. This temporal image sampling constraint is expressed as a function of the navigation trajectory for elementary camera motions. The predicted temporal sampling (also known as frame) rates are on the order of those needed for adaptive optics control systems and require very large bandwidths. The temporal image sampling constraint is then reevaluated by incorporating inertial measurements. The incorporation of inertial measurements is shown to reduce the required temporal sampling rate to practical levels, which evidences the fundamental synergy between image and inertial sensors for navigation and serves as the basis for a real-time adaptive antialiasing strategy.**

***Index Terms*—Image processing, machine vision, navigation.**

I. Introduction

As ORIGINALLY presented in [1], it is well known that optical measurements provide excellent navigation information when properly interpreted. Optical navigation is not new. Pilotage is the oldest and most natively familiar form of navigation to humans and other animals. Mechanical instruments such as astrolabes, sextants, and driftmeters [2] have been used to make precision observations of the sky and ground to improve navigation performance for centuries.

The difficulty in using optical measurements for autonomous navigation, that is, without human intervention, has always been in the interpretation of the image, which is a difficulty shared with automatic target recognition (ATR). Indeed, when celestial observations are used, the ATR problem in this structured environment is tractable, and automatic star trackers are widely used in astro-inertial navigation systems for long-range aircraft, space navigation, and intercontinental ballistic missile guidance. When ground images are to be used, the difficulties associated with image interpretation are paramount. At the same time, the problems associated with the use of optical measurements for navigation are somewhat easier than those of ATR. Moreover, recent developments in feature tracking algorithms, miniaturization, and reduction in cost of inertial sensors and optical imagers, aided by the continuing improvement in microprocessor technology, motivate us to consider using inertial measurements to aid the task of feature tracking in image sequences and realize a tightly coupled image-aided inertial navigation system (INS).

This is an active area of research, and many algorithms exist that attempt to solve this problem by identifying a unique feature in one image and then searching subsequent images for a feature match [3]. This is called the "correspondence search problem." The correspondence problem is plagued by feature ambiguity, temporal feature changes, as well as occlusions, which are difficult for a computer to address. Constraining the correspondence search to a subset of the image plane has the dual advantage of increasing robustness by limiting false matches and improving the search speed. A number of *ad hoc* methods to constrain the correspondence search have been proposed in the literature.

The methods are typically classified as either feature based or optic flow based, depending on how the image correspondence problem is addressed. Feature-based methods determine correspondence for "landmarks" in the scene over multiple frames, while optic-flow-based methods typically determine correspondence for a whole portion of the image between frames using correlation techniques. Optic flow methods have been proposed in the literature, generally for elementary motion detection, in a somewhat structured environment focusing on determining the relative velocity or angular rates for obstacle avoidance [4].

Feature-tracking-based navigation methods have been proposed both for fixed-mount imaging sensors or gimbal-mounted detectors that "stare" at the target of interest; this is similar to the gimballed infrared detector on some heat-seeking missiles. Many feature-tracking-based navigation methods exploit

Manuscript received September 18, 2009; revised January 11, 2010 and May 14, 2010; accepted June 28, 2010. Date of publication July 26, 2010; date of current version October 20, 2010. The review of this paper was coordinated by Dr. K. Deng.

The authors are with the Department of Electrical and Computer Engineering, Air Force Institute of Technology, Wright-Patterson AFB, Dayton, OH 45433 USA (e-mail: michael.veth@afit.edu; richard.martin@afit.edu; meir.pachter@afit.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2010.2060740

knowledge (either *a priori*, through binocular stereopsis, or by exploiting terrain homography) of the target location and solve the inverse trajectory projection problem [5], [6]. If no *a priori* knowledge of the scene is provided, egomotion estimation is completely correlated with estimating the scene. This is referred as the structure-from-motion problem. A theoretical development of the geometry of fixed-target tracking, with no *a priori* knowledge, is provided in [7]. An online (extended-Kalman-filter-based) method for calculating a trajectory by tracking features at an unknown location on the Earth's surface, provided that the topography is known, is given in [8]. Finally, navigation-grade inertial sensors and terrain images collected on a T-38 "Talon" are processed, and the potential benefits of optical-aided inertial sensors are experimentally shown in [9]. The inertially aided feature tracking theory applies to objects with known velocity relative to the camera. The most general class of objects with this property are stationary (relative to the Earth); however, the theory could be applied in other situations where the target velocity might be known. We have not addressed this more-advanced topic in this paper.

Many methods for solving the correspondence problem have been proposed in the computer vision literature. A popular algorithm is the Lucas–Kanade feature tracker [10], which relies on the premise of the invariance of the intensity field between images. It uses a template correlation algorithm to minimize the sum of squared differences between image intensities. The algorithm typically assumes a linear (xy plane) motion model but can be extended to optimize over affine or bilinear transformations. Other feature correspondence algorithms have been proposed that are invariant to rotations, scaling, or both (e.g., [11]). More robust feature tracking algorithms are typically computationally expensive, and a designer must trade tracking robustness and accuracy for real-time performance.

This paper is organized as follows. Current techniques for constraining the correspondence search problem are outlined in Section II. The general spatial-temporal image sampling problem is described in Section III, which provides a theoretical basis for the derivation of a set of sampling constraints in Sections IV and V for situations where inertial aiding is available and unavailable. The effects of relative motion between the camera and the world are derived in Section VI, and the overall theory is demonstrated using an illustrative case study in Section VII.

II. Current Correspondence Constraint Approaches

Exploiting inertial measurements to constrain the correspondence search has been proposed in the literature. In this section, two methods that exploit inertial measurements are discussed.

Bhanu *et al.* [12] utilize inertial measurements to compensate for rotation between images and to predict the focus of expansion in the second image. Once the second image is derotated and the focus of expansion is established, the correspondence between interest points is calculated using goodness-of-fit metrics. One relevant metric is the correspondence search constraint placed on each point. This constraint ensures each interest point lies in a cone-shaped region, with the apex at the focus of expansion, bisected by the line joining the focus of expansion and the interest point in the camera frame at the first image time. While this constraint is not statistically rigorous, it does show the value of using inertial measurements to aid the correspondence problem.

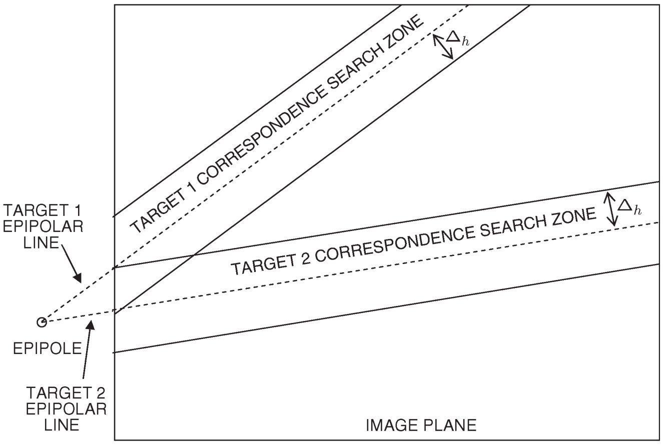


Fig. 1. Correspondence search constraint using epipolar lines. Given a projection of an arbitrary point in an initial image, combined with knowledge of the translation and rotation to a second image, the correspondence search can be constrained to an area near the epipolar line. Note that the epipole can be located outside of the image plane, as shown in this example.

Strelow also incorporates inertial measurements to constrain the correspondence search between image frames [13]. This constraint on the image search space is a similar concept to the field of expansion method proposed by Bhanu *et al.*; however, Strelow generalizes the approach by exploiting epipolar geometry. The projection of an arbitrary point in an image is described by an epipolar line in a second image. All epipolar lines in an image converge at the projection of the focus of the complimentary image. Combining knowledge of the translation and rotation between images and the pixel location of a candidate target in the first image, a correspondence search can then be constrained to an area "near" the epipolar line. This approach is illustrated in Fig. 1.

Strelow's method of using inertial measurements to constrain the correspondence search along an epipolar line is *ad hoc*, since the search space is not statistically defined. This method could be improved by utilizing a stochastically rigorous development.

In previous publications, we have presented an approach that leverages the inertial measurements and any available terrain information to predict the locations and statistical uncertainty of features in a new image [14], [15]. Our goal in this paper is to expand the stochastic constraint theory to an elemental level that is dependent on the inherent optical properties of the sensor. Analyzing the correspondence problem from this perspective reveals the parallel nature between feature correspondence searching and temporal sampling theory in signal processing, which is well understood. As a result, feature correspondence ambiguity is shown to be analogous to temporal aliasing. Thereby, sampling theory can be used to predict and mitigate/avoid the presence of temporal aliasing in the feature space.

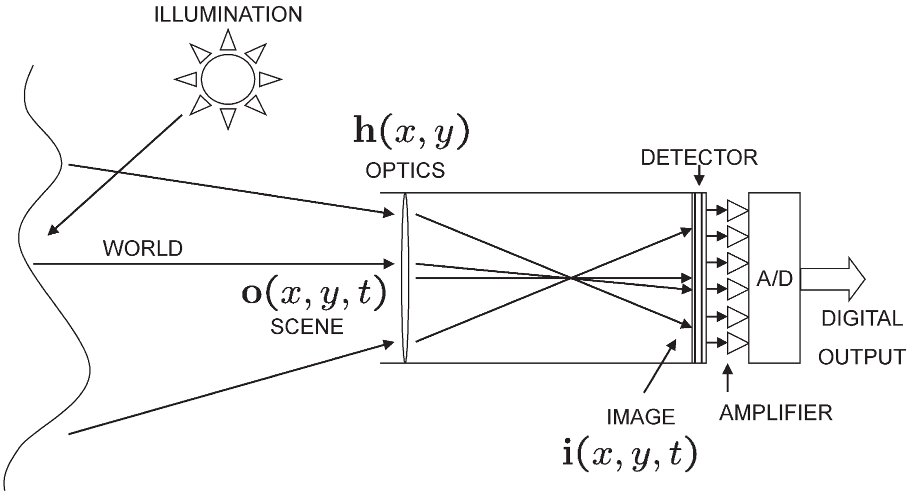


Fig. 2. Digital imaging system. The imaging system transforms the scene into a digital image. The major components of the camera are the optics, the light detector, the amplifier, and the analog-to-digital converter.

In the next section, the theory of image sampling is developed from first principles, with particular attention paid to the anticipated issues with regard to temporal sampling.

III. General Image Sampling Problem

The mathematical relationships governing spatial-temporal sampling are developed from basic optical and sampling theory. This development provides a theoretical basis that is used to develop temporal sampling constraints in subsequent sections. In this paper, temporal sampling rates and frame rates are used synonymously.

Notation: $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $\mathrm{E}\{\cdot\}$ denote complex conjugate, matrix transpose, Hermitian (conjugate) transpose, and statistical expectation, respectively, and $\mathrm{j} = \sqrt{-1}$ is the unit imaginary number. Throughout, ν denotes a spatial frequency in units of *per meter*, and f denotes a temporal frequency in units of *hertz*.

For clarity, the applicable sensor modeling development is included from [1]. A digital imaging device is, in essence, a sampler of light intensity patterns in three dimensions: two spatial and one temporal. Analyzing the effects of the sampling process on image sequences resulting from camera motion with due regard given to the motion's dynamics has very important implications on how to properly interpret image sequences to derive navigation information.

A. Effects of Egomotion on Image Formation

As discussed in the previous section, the recorded image is a representation of the optical intensity patterns generated by a scene. The projection function is a function of the scene itself, the camera optical properties, and the pose (i.e., relative position and orientation) of the camera and scene. This strong coupling between camera pose and the image is the basis for the rapidly growing research efforts dedicated to exploiting images to determine changes in camera pose. In this section, the geometric projection function is developed using a pinhole camera model. This model will be used as a basis to quantify the effects of egomotion and temporal sampling.

B. Optical Sensor Model

An optical sensor is a device designed to measure the intensity of optical energy (light) entering the sensor through an aperture. Imaging sensors consist of an array of light-sensitive detectors that create a 2-D light intensity measurement (i.e., sampled image). In this section, the basic physical properties of an optical sensor are presented, and a model representing an optical sensor is given.

For the purposes of this discussion, the *world* is defined as a collection of all real objects. Some objects are sources of radiometric illumination or *radiance*. These light sources illuminate the world and interact with the other physical objects through various types of reflection. The amount of light along a certain direction is defined as the *irradiance* [3]. The physical irradiance pattern entering the aperture of the optical sensor is defined as the *scene* and is represented by a continuous array of nonnegative real numbers, i.e., $\mathbf{o}(x,y,t)$, projected onto the image plane. For the purposes of this discussion, the irradiance sources are constrained to an arbitrary piecewise-continuous Lambertian surface in three dimensions.

A digital optical imaging sensor consists of an aperture, a lens, a detector array, and a sampling array, as shown in Fig. 2. The lens focuses the scene on the detector array. The light pattern focused on the detector array is defined as the *image* and is represented by $\mathbf{i}(x,y,t)$. In statistical terms, the *image* is the mean photon arrival rate and is defined by a Poisson distribution [16]. The detector array converts the light energy into a voltage or a charge that is converted to a digital value by the sampling array. The sampling array is assumed to be a square grid, although other patterns can be designed (e.g., honeycomb) [17].

The lens is an analog low-pass filter in the spatial domain, with a cutoff frequency (ν_c) determined by the aperture (D), wavelength of light source (λ), and focal length of the camera (f_0) [16]:

$$\nu_c = \frac{D}{\lambda f_0} \tag{1}$$

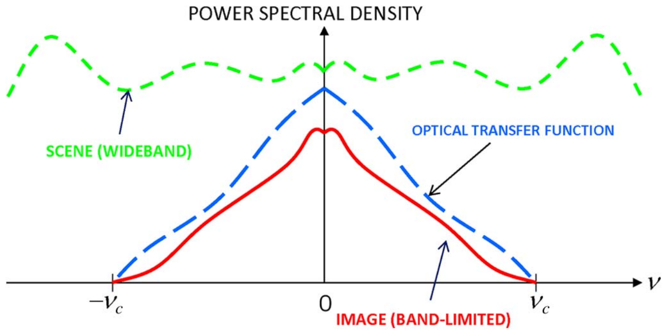


Fig. 3. Effects of camera optics on image spatial frequency. The camera optics act as a low-pass filter with a cutoff frequency of ν_c. The scene, which is wideband, appears as a band-limited image on the detector array.

Thus, a scene consisting of a point source of light (delta function intensity) would appear slightly blurred (spread) on the image plane. Assuming spatial invariance, this blurring due to the lens is represented by the *point spread function* (PSF), which is denoted as $\mathbf{h}(\xi, \rho)$, where ξ and ρ are the spatial differences in the x- and y-directions, respectively. The image in the spatial domain can now be mathematically expressed as the convolution of the scene and PSF [18], i.e.,

$$\mathbf{i}(x, y, t) = \int\limits_{\xi \in \mathbf{X}} \int\limits_{\rho \in \mathbf{Y}} \mathbf{o}(\xi, \rho, t)\mathbf{h}(x - \xi, y - \rho)\, d\rho\, d\xi. \tag{2}$$

The image is physically continuous in space and time. This continuous function of three variables is then sampled and converted to an array of (digital) numbers. Concerning the sample process, the light energy in the image is integrated in each pixel over a temporal period defined as the *dwell time* (Δt). The sampled image ($\mathbf{i}_s[m, n, t_k]$) is obtained for integer pixel location (m, n) and sample time t_k as

$$\mathbf{i}_s(m, n, t_k) = \int\limits_{m-\Delta x/2}^{m+\Delta x/2} \int\limits_{n-\Delta_y/2}^{n+\Delta_y/2} \int\limits_{t_k-\Delta t/2}^{t_k+\Delta t/2} \mathbf{i}(x, y, t_k)\, dx\, dy\, dt \tag{3}$$

where Δx and Δy correspond to the pixel dimensions in the x- and y-directions, respectively. The frequency-domain representation of the PSF $\mathbf{H}(\nu_x, \nu_y)$ is called the *optical transfer function*. Applying the Fourier transform to the image equation (2) yields

$$\mathbf{i}(\nu_x, \nu_y, t) = \mathbf{o}(\nu_x, \nu_y, t)\mathbf{h}(\nu_x, \nu_y). \tag{4}$$

In most conditions, the projected scene can be treated as a wideband function relative to the optical transfer function (i.e., $\nu_{c_{\text{scene}}} \gg \nu_{c_{\text{OTF}}}$). This results in the following spatial frequency limitation of the projected image:

$$\mathbf{i}(\nu_x, \nu_y, t) = 0 \qquad \forall \sqrt{\nu_x^2 + \nu_y^2} > \nu_c \tag{5}$$

with autocorrelation function denoted as $R_{\mathbf{i}}(x, y)$. This relationship is graphically expressed in Fig. 3.

For simplicity, we will assume that the primary time dependence of the observations is due to spatial translations, i.e.,

$$\mathbf{i}(x, y, t) = \mathbf{i}\left(x - \delta_x(t), y - \delta_y(t), 0\right). \tag{6}$$

This is a reasonable assumption if we are rapidly sampling and focusing on a small subregion of the image, e.g., for feature tracking. Due to the various effects discussed in [1], we assume that the instantaneous velocity of the translation is $(\dot{s}_x^{\text{proj}}, \dot{s}_y^{\text{proj}})$, which is bounded in magnitude by $\dot{s}_{\max}$.

The cutoff frequency of the PSF dictates that the image be spatially sampled at

$$\Delta_x = \Delta_y = \Delta_{\text{pixel}} \leq \frac{1}{2\nu_c}. \tag{7}$$

The goal of this paper is to determine the temporal sampling rate necessary to enable feature correspondence determination across successive frames with minimal effort. This is equivalent to requiring that a feature not move more than one pixel between successive frames. This, in turn, is equivalent to requiring that the image be sampled fast enough such that there is no temporal aliasing in a given pixel when viewed as a function of time.

IV. Inertial Navigation System-Unaided Temporal Sampling Requirements

Given the model in Section III-B, the worst-case scenario is when the scene is being translated with a constant rate $\dot{s}_{\max}$ in either direction. Assuming that the motion is in the x-direction, viewing pixel (x_o, y_o) over time yields the continuous-time intensity signal $u(t)$ given by

$$u(t) \triangleq \mathbf{i}(x_o, y_o, t) = \mathbf{i}(x_o + \dot{s}_{\max}t, y_o, 0). \tag{8}$$

The Fourier transform of $u(t)$ is related to the PSF, which, in turn, will provide a Nyquist criterion for sampling. It is easy to derive the 1-D Fourier transform pair, i.e.,

$$\mathbf{i}(x, y_o, 0) \Longleftrightarrow \int \mathbf{I}(\nu, \nu_y)e^{\mathrm{j}2\pi\nu_y y_o}\, d\nu_y \triangleq V(\nu). \tag{9}$$

Since $\mathbf{I}(\nu_x, \nu_y) = 0$ for $\nu_x \geq \nu_c$, we also have $V(\nu) = 0$ for $\nu \geq \nu_c$. Using $V(\nu)$, the Fourier transform of $u(t)$ is

$$U(f) = \int \mathbf{i}(x_o + \dot{s}_{\max}t, y_o, 0)e^{-\mathrm{j}2\pi ft}\, dt \tag{10}$$

$$= \frac{1}{\dot{s}_{\max}}e^{\mathrm{j}2\pi(f/\dot{s}_{\max})x_o} \int \mathbf{i}(x, y_o, 0)e^{-\mathrm{j}2\pi(f/\dot{s}_{\max})x}\, dx \tag{11}$$

$$= \frac{1}{\dot{s}_{\max}}e^{\mathrm{j}2\pi(f/\dot{s}_{\max})x_o}V(f/\dot{s}_{\max}) \tag{12}$$

and $U(f) = 0$ for $f/\dot{s}_{\max} \geq \nu_c$ or, equivalently, for $f \geq \dot{s}_{\max}\nu_c$. Thus, the Nyquist temporal sampling criterion is

$$f_t \geq 2\dot{s}_{\max}\nu_c. \tag{13}$$

Assuming square pixels that are sized according to the spatial Nyquist sampling constraint (i.e., $\nu_x = \nu_y = 2\nu_c$) results in the following pixel size (7):

$$\Delta_{\text{pixel}} = \frac{1}{2\nu_c}. \tag{14}$$

Substituting (14) into (13) results in the normalized temporal sampling constraint, i.e.,

$$f_t \geq \frac{\sqrt{\left(\dot{s}_x^{\mathrm{proj}}\right)^2 + \left(\dot{s}_y^{\mathrm{proj}}\right)^2}}{\Delta_{\mathrm{pixel}}}. \tag{15}$$

As a result, to minimize temporal aliasing, the Nyquist rate can be achieved by ensuring that no feature moves more than one half of the minimum distance between intensity peaks in the image plane. Given an optical cutoff frequency of ν_c, the temporal sampling interval T_s should be chosen such that the maximum image shift due to camera motion is less than $1/2\nu_c$. This implies a fundamental interrelationship between the minimum spatial and temporal sampling intervals, which is somewhat similar to the spatial-temporal discretization constraint found when solving the heat partial differential equation, which is also known as the von Neumann condition [19].

In the next section, a mathematical model describing the relationship between point locations in the world and the image will be derived. The resulting projection equations will be used to calculate appropriate temporal sampling intervals, based on scene geometry and camera motion.

V. Inertial Navigation System-Aided Sampling

Now, assume that an INS is used to recenter each frame based on the estimated motion of the camera with respect to the scene. The time-dependent translation becomes a random process that is represented by $\delta_x(t)$ and $\delta_y(t)$ in the x- and y-directions, respectively. For simplicity, we model this as a Wiener process, with distribution

$$\begin{bmatrix} \delta_x(t) \\ \delta_y(t) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 t & 0 \\ 0 & \sigma^2 t \end{bmatrix}\right) \tag{16}$$

where σ is given in [1] as

$$\sigma = f_0\sqrt{q_w} \tag{17}$$

$$q_w = 4.2 \times 10^{-7} \tag{18}$$

where σ is in units of $m\sqrt{s}$, and q_w is in units of rad/s.

To determine the sampling rate, we must find the cutoff frequency of the power spectral density (PSD) of the random process, i.e.,

$$\mathbf{i}(x_o, y_o, t) = \mathbf{i}\left(x_o - \delta_x(t), y_o - \delta_y(t), 0\right). \tag{19}$$

For simplicity, consider a 1-D version of the problem, i.e.,

$$\mathbf{i}(x_o, t) = \mathbf{i}\left(x_o - \delta(t), 0\right). \tag{20}$$

The PSD of $\mathbf{i}(x_o, t)$ is the Fourier transform of its autocorrelation function. The autocorrelation function is

$$R(\tau) = \mathrm{E}\left\{\mathbf{i}(x_o, 0)\mathbf{i}\left(x_o - \delta\left(|\tau|\right), 0\right)\right\}. \tag{21}$$

Treating $\mathbf{o}$ as a random process and breaking up the expectation into expectation over $\mathbf{o}$ and δ in turn, we get the following:

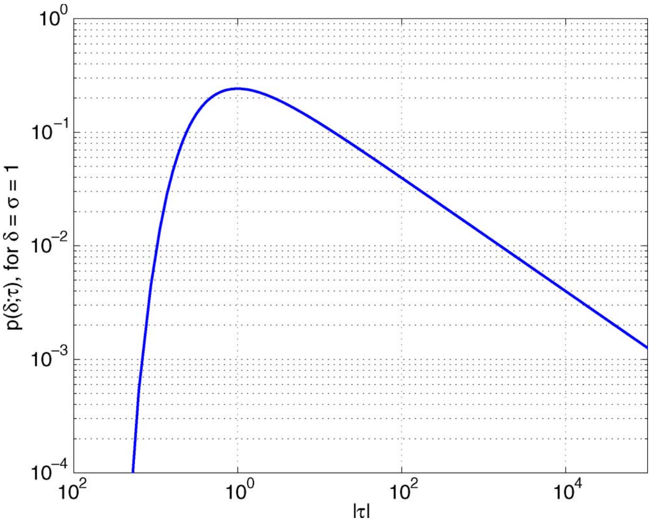


Fig. 4. Function $p_\delta(\delta;\tau)$ for $\delta = \sigma = 1$, as a function of τ.

$$\begin{aligned} R(\tau) &= \mathrm{E}_{(\delta\tau)}\left\{\mathrm{E}_{\mathbf{o}}\left\{\mathbf{i}(x_o, 0)\mathbf{i}\left(x_o - \delta\left(|\tau|\right), 0\right)\right\}\right\} && (22)\\ &= \int p_\delta(\delta\tau) \underbrace{\left[\int p_{\mathbf{o}}(\mathbf{o})\mathbf{i}(x_o, 0)\mathbf{i}(x_o - \delta, 0)d\mathbf{o}\right]}_{R_{\mathbf{i}}(\delta)} d\delta \\ &= \mathrm{E}_{(\delta\tau)}\left\{R_{\mathbf{i}}(\delta)\right\}. && (23) \end{aligned}$$

Thus, the autocorrelation function of a pixel is the expected value of the autocorrelation of a single frame of the image. The expected value is taken with respect to a spatial shift of δ, where δ is normally distributed as in (16), and the time variable τ is a parameter of the distribution $p_\delta(\delta;\tau)$. This distribution is shown as a function of τ in Fig. 4.

Taking the Fourier transform of $R(\tau)$ with respect to τ and then representing $R_{\mathbf{i}}(\delta)$ in terms of its Fourier transform, we get the following:

$$\begin{aligned} S(f) &= \int \left[\int p_\delta(\delta\tau)R_{\mathbf{i}}(\delta)d\delta\right] e^{-\mathrm{j}2\pi f\tau}\, d\tau \\ &= \int\int p_\delta(\delta\tau)\int S_{\mathbf{i}}(\nu)e^{\mathrm{j}2\pi\nu\delta}d\nu d\delta e^{-\mathrm{j}2\pi f\tau}\, d\tau \\ &= \int\int S_{\mathbf{i}}(\nu)\left[\int p_\delta(\delta\tau)e^{\mathrm{j}2\pi\nu\delta}d\delta\right] e^{-\mathrm{j}2\pi f\tau}\, d\nu\, d\tau. && (24) \end{aligned}$$

The bracketed term in the last line of (24) is the Fourier transform of a Gaussian, which is itself a Gaussian; hence

$$S(f) = \int S_{\mathbf{i}}(\nu)\left[\int e^{-(2\pi^2\sigma^2\nu^2)|\tau|}e^{-\mathrm{j}2\pi f\tau}\, d\tau\right] d\nu. \tag{25}$$

The bracketed term in (25) is the Fourier transform of a two-sided exponential, which can also be computed via a lookup table, i.e.,

$$S(f) = \int S_{\mathbf{i}}(\nu)\left[\frac{\sigma^2\nu^2}{\pi^2\sigma^4\nu^4 + f^2}\right] d\nu. \tag{26}$$

If a similar analysis is performed in 2-D, the result analogous to (26) can be shown to be

$$S(f) = \int\int S_{\mathbf{i}}(\mu,\nu)\left[\frac{\sigma^2(\mu^2+\nu^2)}{\pi^2\sigma^4(\mu^2+\nu^2)^2+f^2}\right]d\mu\,d\nu. \quad (27)$$

Equation (27) cannot be further simplified without assuming a specific form for $R_{\mathbf{i}}(\delta_x,\delta_y)$ or, equivalently, for $S_{\mathbf{i}}(\mu,\nu)$. However, note that we have previously assumed that $S_{\mathbf{i}}(\mu,\nu) = 0$ for $\sqrt{\mu^2+\nu^2} \geq \nu_c$. Moreover, the bracketed term in (27) will rapidly drop off for $f > \pi\sigma^2(\mu^2+\nu^2)$. Thus, (27) has an approximate cutoff frequency of $\pi\sigma^2\nu_c^2$, yielding an approximate Nyquist temporal sampling criterion of

$$\begin{aligned} f_t \gtrsim 2\pi\sigma^2\nu_c^2 \mid_{\nu_c=\frac{1}{2\Delta_{\text{pixel}}}} &= 2\pi f_0^2 q_w\left(\frac{1}{4\Delta_{\text{pixel}}^2}\right) \\ &= \frac{\pi}{2}\frac{f_0^2 q_w}{\Delta_{\text{pixel}}^2}. \end{aligned} \quad (28)$$

Thus, a rough bound for the maximum sampling interval is

$$T_s \leq \left(\frac{2}{\pi}\right)\frac{\Delta_{\text{pixel}}^2}{f_0^2 q_w}. \quad (29)$$

The rate of variance increase σ^2 has units of square meters per second and the frequency ν_c has units of per meter; hence, f_t has units of hertz.

Consider the specific cases of a PSD with a triangular or a rectangular cross section, i.e.,

$$S_{\mathbf{i}}^{\text{tri}}(\mu,\nu) = \begin{cases} 1-\sqrt{\mu^2+\nu^2}/\nu_c, & \sqrt{\mu^2+\nu^2} \leq \nu_c \\ 0, & \text{else} \end{cases} \quad (30)$$

$$S_{\mathbf{i}}^{\text{rect}}(\mu,\nu) = \begin{cases} 1, & \sqrt{\mu^2+\nu^2} \leq \nu_c \\ 0, & \text{else.} \end{cases} \quad (31)$$

The triangular PSD is representative of most imaging devices. On the other hand, the rectangular PSD will give us a bound on performance and force us to choose the most conservative sampling frequency, since it has the most high-frequency content of all PSDs with a cutoff frequency of ν_c. Converting (27) into polar coordinates, we get the following:

$$S(f) = \int_0^{2\pi}\int_0^{\nu_c} S_{\mathbf{i}}(\rho,\theta)\left[\frac{\sigma^2\rho^2}{\pi^2\sigma^4\rho^4+f^2}\right]\rho\,d\rho\,d\theta. \quad (32)$$

Making the substitutions $z = \sigma\rho/\sqrt{f}$ and $z_c = \sigma\nu_c/\sqrt{f}$ and substituting in the two example PSDs, we get the following:

$$S^{\text{tri}}(f) = \frac{2\pi}{\sigma^2}\int_0^{z_c}\left(1-\frac{z}{z_c}\right)\frac{z^3}{\pi^2 z^4+1}dz \quad (33)$$

$$S^{\text{rect}}(f) = \frac{2\pi}{\sigma^2}\int_0^{z_c}\frac{z^3}{\pi^2 z^4+1}dz. \quad (34)$$

The first integral can numerically be evaluated as a function of z_c, allowing numerical evaluation of $S^{\text{tri}}(f)$ in terms of $f/(\pi\sigma^2\nu_c^2)$. The latter integral can be evaluated in closed form as

$$S^{\text{rect}}(f) = \frac{1}{2\pi\sigma^2}\ln\left(1+\frac{\pi^2\sigma^4\nu_c^4}{f^2}\right). \quad (35)$$

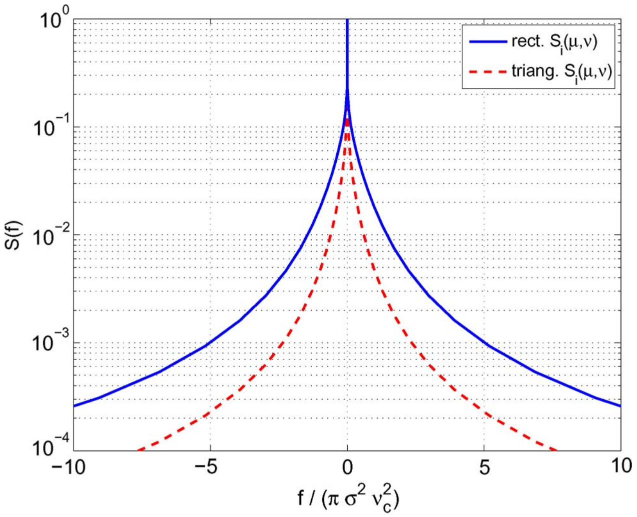


Fig. 5. PSD of a pixel over time, recentered using the INS measurements, assuming a triangular or a rectangular PSD for a single observed image frame.

The two resulting PSDs are shown in Fig. 5. Note that they exhibit a delta-function-like behavior in that, at $f = 0$, they go to infinity, but the spikes are infinitesimally narrow with a finite area.

To obtain an approximate temporal Nyquist sampling criterion, consider the fraction of the energy in each PSD within $-\Delta < f < \Delta$. For the triangular image PSD, this must be numerically evaluated, but for the rectangular image PSD, it can be shown that

$$\frac{\int_{-\Delta}^{\Delta} S^{\text{rect}}(f)df}{\int_{-\infty}^{\infty} S^{\text{rect}}(f)df} = \frac{2}{\pi}\tan^{-1}(\gamma)+\frac{\gamma}{\pi}\ln(1+\gamma^{-2}) \quad (36)$$

$$\gamma \triangleq \frac{\Delta}{\pi\sigma^2\nu_c^2}. \quad (37)$$

One minus the ratio in (36) and the analogous numerical result for the triangular PSD are plotted in Fig. 6. The approximate bound in (28), corresponding to $\gamma = 1$, captures all but 30% of the energy for the conservative bound (from the rectangular image PSD) and all but 20% of the energy of the more representative bound (from the triangular image PSD). The amount of energy that causes temporal aliasing will drop to 2% to 3% if the sampling frequency is increased by an order of magnitude.

VI. Egomotion Effects on Temporal Sampling

In the previous section, the effects of egomotion on the formation of the image are presented. In this section, the egomotion effects on temporal sampling are illustrated. Reducing the spatial dimensionality of the problem from two to one is performed to illustrate the effects of egomotion on temporal sampling in a manner that is easier to visualize.

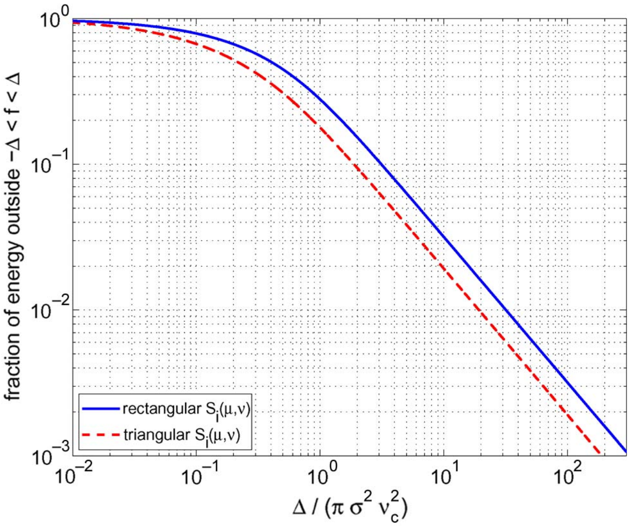


Fig. 6. Fraction of energy of the PSD outside of $-\Delta < f < \Delta$. The solid curve was analytically evaluated, and the dashed curve was numerically evaluated.

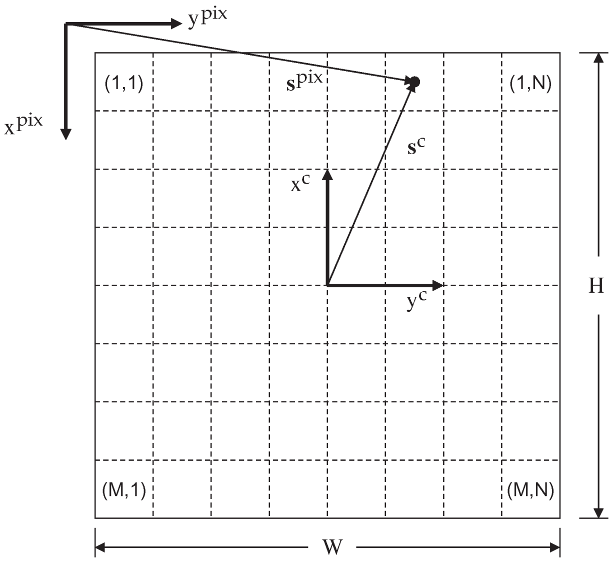


Fig. 7. Camera image array. The camera imager consists of an $(M \times N)$ array of pixels. The physical height and width of the array are represented by H and W, respectively.

A. Reference Frames and Notation

In this paper, three reference frames are used to describe the general imaging problem. The *navigation reference frame* is a 3-D orthogonal basis used to represent locations relative to the world with respect to an arbitrary origin. A vector coordinitized in the navigation reference frame is denoted using the "n" superscript, e.g., $\mathbf{p}^n$. The *camera reference frame* is a 3-D orthogonal basis with origin located at the optical center of the camera, with the z-axis pointing toward the principal point (out the front of the camera). The x- and y-axes are aligned as shown in Fig. 7. A vector coordinitized in the camera reference frame is denoted using the "c" superscript, e.g., $\mathbf{s}^c$. Finally, the *pixel plane* reference frame is a 2-D orthogonal basis used to represent a pixel location on the image plane. A vector coordinitized in the pixel plane reference frame is denoted using the "pix" superscript, e.g., $\mathbf{s}^{pix}$. The pixel plane reference frame is measured in units of pixels. For simplicity, it is assumed that the inertial sensor is rigidly mounted to the imaging device such that motion of the camera with respect to the inertial frame is observed. This assumption can be made without loss of generality as *any* camera motions are captured whether the camera/IMU sensor is mounted on a pan/tilt rig or when mounted on a 6-degree-of-freedom platform.

B. Projection Theory

The camera optical properties define the relationship between the scene and the projected image. Recalling the simple camera model (see Fig. 2), the lens focuses the incoming irradiance pattern (i.e., scene) onto the image plane. For a theoretical thin lens, the projection is a function of the focal length of the lens and the distance from the lens. This relationship is expressed by the *fundamental equation of the thin lens* [3], i.e.,

$$\frac{1}{Z} + \frac{1}{z} = \frac{1}{f_0} \tag{38}$$

where Z is the distance from the object to the lens, z is the distance from the lens to the image plane, and f_0 is the focal length.

As the aperture of the thin lens decreases to zero, the system can be modeled as a pinhole camera. In this model, all incoming light must pass through the optical center and is projected on an image plane located at a distance f from the lens. The resulting image is an inverted projection of the scene.

This model can further be simplified by placing a virtual image plane in front of the optical center. Given a point source at location $\mathbf{s}^c \in \Re^3$, the resulting location of the point source on the image plane, relative to the optical center of the camera, is given by

$$\mathbf{s}^c_{\text{proj}} = \left(\frac{f_0}{s^c_z}\right)\mathbf{s}^c = f_0\underline{\mathbf{s}}^c \tag{39}$$

where s^c_z is the distance of the point source from the optical center of the camera in the z_c direction. The underline indicates a vector expressed in homogeneous notation, which is given by

$$\underline{\mathbf{s}}^c = \frac{1}{s^c_z}\mathbf{s}^c. \tag{40}$$

To interpret the calculated projection in a digital image, the physical image plane coordinates must be converted to a coordinate system based on pixel location. The following development defines the pixel coordinate system and derives the transformation from the physical image plane to the pixel location. The image plane consists of an $(M \times N)$ grid of rectangular pixels with height H and width W, shown in Fig. 7. The origin of the projection frame is located at the physical center of the array. The origin of the pixel coordinate system is located beyond the upper left corner of the array, such that the center of the upper left pixel corresponds to the (1, 1) pixel coordinate. This definition of pixel coordinates corresponds to

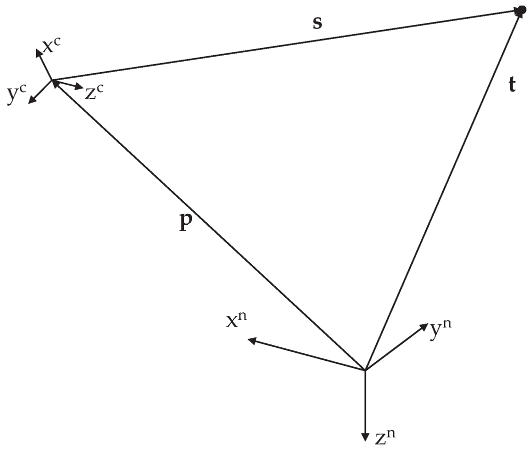


Fig. 8. Target to image transformation geometry. The relationship between the camera position $(\mathbf{p})$ and the target location $(\mathbf{t})$ can be expressed in pixel coordinates using transformations based on the navigation state and intrinsic camera parameters.

the elemental matrix locations when the image is stored in a computer. This can be expressed as a two-element vector, i.e.,

$$\mathbf{s}^{\text{pix}} = \begin{bmatrix} u \\ v \end{bmatrix} \tag{41}$$

where u and v are the row and column corresponding to the pixel of interest, with units of *pixels*.

The transformation from the projection coordinates to pixel coordinates is given by

$$\mathbf{s}^{\text{pix}} = \begin{bmatrix} -\frac{1}{\Delta_x} & 0 & 0 \\ 0 & \frac{1}{\Delta_y} & 0 \end{bmatrix} \mathbf{s}^c_{\text{proj}} + \begin{bmatrix} \frac{M+1}{2} \\ \frac{N+1}{2} \end{bmatrix} \tag{42}$$

where Δ_x and Δ_y are the sizes of the pixels in the x-and y-directions, respectively, which are defined as

$$\Delta_x = \frac{H}{M} \tag{43}$$

$$\Delta_y = \frac{W}{N}. \tag{44}$$

Combining (39) and (42) and expressing the projected pixel location vector using homogeneous coordinates yields the following affine transformation from the camera frame to the pixel location:

$$\mathbf{s}^{\text{pix}} = \begin{bmatrix} -\frac{f_0}{\Delta_x} & 0 & \frac{M+1}{2} \\ 0 & \frac{f_0}{\Delta_y} & \frac{N+1}{2} \end{bmatrix} \underline{\mathbf{s}}^c \tag{45}$$

$$= \mathbf{T}^{\text{pix}}_c \underline{\mathbf{s}}^c. \tag{46}$$

A transformation from a landmark location in navigation frame coordinates to pixel coordinates can now be derived based on the navigation state. The geometry is shown in Fig. 8. The line-of-sight vector $\mathbf{s}$ is the vector difference between the target location $\mathbf{t}$ and the camera position, which are both available in navigation frame coordinates, i.e.,

$$\mathbf{s}^n = \mathbf{t}^n - \mathbf{p}^n. \tag{47}$$

The resultant vector can be transformed to the camera reference frame using the navigation-to-camera frame direction cosine matrix, i.e.,

$$\mathbf{s}^c = \mathbf{C}^c_n \mathbf{s}^n. \tag{48}$$

Finally, the pixel location is calculated using (46).

C. Apparent Pixel Motion Calculations

The previous development is extended to illustrate the apparent pixel motion of a point feature due to relative motion. The development begins by recalling the camera-to-pixel transformation shown in (39)–(48), i.e.,

$$\mathbf{s}^{\text{pix}} = \mathbf{T}^{\text{pix}}_c \underline{\mathbf{s}}^c = \mathbf{T}^{\text{pix}}_c \mathbf{s}^c / s^c_z \tag{49}$$

where the camera frame line of sight vector $\mathbf{s}^c$ is given by

$$\mathbf{s}^c = \mathbf{C}^c_n [\mathbf{t}^n - \mathbf{p}^n]. \tag{50}$$

The apparent pixel motion is derived by taking the derivative of $\mathbf{s}^{\text{pix}}$ with respect to time, i.e.,

$$\dot{\mathbf{s}}^{\text{pix}} = \mathbf{T}^{\text{pix}}_c \dot{\underline{\mathbf{s}}}^c \tag{51}$$

where

$$\dot{\underline{\mathbf{s}}}^c = \frac{s^c_z \dot{\mathbf{s}}^c - \mathbf{s}^c \dot{s}^c_z}{\left(s^c_z\right)^2}. \tag{52}$$

The time derivative of the camera frame line-of-sight vector is given by

$$\dot{\mathbf{s}}^c = \mathbf{C}^c_n \mathbf{\Omega}^n_{cn} [\mathbf{t}^n - \mathbf{p}^n] + \mathbf{C}^c_n [\dot{\mathbf{t}}^n - \dot{\mathbf{p}}^n] \tag{53}$$

where $\mathbf{\Omega}^n_{cn}$ is the skew-symmetric form of the angular rate of the camera to the navigation frame, which is expressed in the navigation frame. The skew-symmetric form is defined in [20]. Expressing the rotations in the camera frame yields the following equivalent form:

$$\dot{\mathbf{s}}^c = -\mathbf{\Omega}^c_{nc} \mathbf{s}^c + \mathbf{C}^c_n [\dot{\mathbf{t}}^n - \dot{\mathbf{p}}^n]. \tag{54}$$

An analysis of (54) shows that the change in line-of-sight vector is a function of both the camera rotation and relative translational motion between the camera and landmark of interest.

In many cases, the landmark motion relative to the navigation frame is insignificant and can be neglected. Applying this assumption and coordinitizing the camera translational motion in the camera frame yields

$$\dot{\mathbf{s}}^c = -\mathbf{\Omega}^c_{nc} \mathbf{s}^c - \mathbf{v}^c \tag{55}$$

where $\mathbf{v}^c$ is the velocity of the camera, relative to the navigation frame, coordinitized in the camera frame. Combining (51), (52), and (55) results in the well-known *optical flow equations* [21], which are the apparent motion in the x- and y-directions, i.e.,

$$\dot{u} = -\frac{f_0}{\Delta x}\left(-\omega_y - \frac{v_x}{s_z^c} + \frac{s_x^c s_y^c}{\left(s_z^c\right)^2}\omega_x - \left(\frac{s_x^c}{s_z^c}\right)^2 \omega_y + \frac{s_y^c}{s_z^c}\omega_z + \frac{s_x^c}{\left(s_z^c\right)^2} v_z\right) \tag{56}$$

$$\dot{v} = \frac{f_0}{\Delta_y}\left(\omega_x - \frac{v_y}{s_z^c} - \frac{s_x^c s_y^c}{\left(s_z^c\right)^2}\omega_y + \left(\frac{s_y^c}{s_z^c}\right)^2 \omega_x - \frac{s_x^c}{s_z^c}\omega_z + \frac{s_y^c}{\left(s_z^c\right)^2} v_z\right) \tag{57}$$

which is expressed using the scalar components of the rotation, velocity, and line-of-sight vectors referenced in (55).

The temporal sampling constraint proposed in the previous section indicates that it is desirable to sample such that the apparent pixel motion is limited to no more than one pixel of change per image in both the x and y spatial dimensions, provided that the image is sampled at spatial Nyquist frequency. Given a sample interval T_s, the maximum pixel motion component $K_{\max}$ can be approximated by

$$K_{\max} = \sqrt{\dot{u}^2 + \dot{v}^2}\, T_s \tag{58}$$

which is subject to the following constraint:

$$K_{\max} \leq 1 \tag{59}$$

to guarantee nonaliased tracking.

In the next section, the derived apparent pixel motion is analyzed for a representative scenario that illustrates the difficulty in achieving samples from traditional imaging systems that do not violate the temporal sampling constraints presented above. The key idea is to determine the required frame rates required to eliminate the effects of *temporal* aliasing, assuming that each image is spatially sampled at or above the Nyquist spatial frequency. It will be shown that to ensure temporal sampling constraints, either relatively high frame rates will be required or antitemporal aliasing filtering will be required. Finally, we propose a solution for eliminating temporal aliasing by incorporating IMU data.

VII. Illustrative Case Study

In this section, the apparent pixel motion is calculated for a selection of representative imaging scenarios. As previously developed, the generalized sampling characteristics of a given imaging sensor are a function of a number of parameters. In this scenario, we will assume that the camera intrinsic parameters (i.e., Δ_x, Δ_y, f_0, D, and λ) are fixed in such a way to guarantee proper spatial sampling. For this case, we are interested in the resulting temporal sampling rate (f_t) that is consistent with the temporal sampling constraints derived in the previous section. Note that changing the frame rate will not affect the *resolution* of the images; however, it will affect the rate of change of each pixel over time, which directly affects the presence of temporal aliasing. The camera intrinsic parameters are chosen to be representative of currently available machine vision cameras. These parameters are shown in Table I.

TABLE I
Camera Intrinsic Parameters. The Camera Intrinsic Parameters Are Chosen to be Representative of Currently Available Machine Vision Cameras and Are Chosen to Eliminate Spatial Aliasing

Description	Parameter	Value	(Units)
Wavelength	λ	550	μm
Focal length	f_0	6	mm
Lens Aperture	D	6/16	mm
Vertical Image Size	M	1024	$pixels$
Vertical Pixel Size	Δ_x	4.4	μm
Horizontal Image Size	N	1280	$pixels$
Horizontal Pixel Size	Δ_y	4.4	μm

The first case study is a simple 5°/s horizontal pan, with no translational motion. The resulting motion parameters for this condition are given as follows:

$$\mathbf{v}^c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \left(\frac{\mathrm{m}}{\mathrm{s}}\right) \tag{60}$$

$$\boldsymbol{\omega}_{nc}^c = \begin{bmatrix} 5\frac{\pi}{180} \\ 0 \\ 0 \end{bmatrix} \left(\frac{\mathrm{rad}}{\mathrm{s}}\right). \tag{61}$$

Substituting these motion parameters and the intrinsic camera parameters into (56) and (57) yields

$$K_{\max} = \frac{6\ \mathrm{mm}}{4.4\ \mu\mathrm{m}} T_s \left\{ \left(\frac{s_x^c s_y^c}{\left(s_z^c\right)^2}\right)^2 + \left[1 + \left(\frac{s_y^c}{s_z^c}\right)^2\right]^2 \right\}^{1/2} \frac{5\pi}{180}. \tag{62}$$

As evinced in (62), the pixel motion is primarily a function of the camera motion with second-order effects related to the position of the point source within the image. The worst-case condition occurs at the extreme extents of the image. Substituting these conditions into (62) yields

$$K_{\max} = \frac{6\ \mathrm{mm}}{4.4\ \mu\mathrm{m}} T_s \{1.233\} \frac{5\pi}{180} \tag{63}$$

$$= 146.7 T_s. \tag{64}$$

Applying the temporal sampling constraint and solving for T_s yields

$$T_s \leq \frac{1}{146.7} \quad (\mathrm{s}) \tag{65}$$

which results in a minimum frame rate of 146.7 Hz and, consequently, a maximum exposure time of 6.8 ms.

As shown in (56) and (57), the relationship between pan and tilt rates and the required sampling rate is linear. In addition, the relationship between the roll rate and the required sampling rate is linear and is strongly related to the location of the object in the image (e.g., objects far from the center of rotation have greater apparent motion).

In the next example, the effects of translational motion are investigated. Here, the camera is moving at 300 m/s with a fixed orientation. The distance to the terrain is 10 000 m, which represents a high-altitude cruise profile for an aircraft. The resulting motion parameters for this condition are given as follows:

$$\mathbf{v}^c = \begin{bmatrix} 300 \\ 0 \\ 0 \end{bmatrix} \left(\frac{\text{m}}{\text{s}}\right) \tag{66}$$

$$\boldsymbol{\omega}_{nc}^c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \left(\frac{\text{rad}}{\text{s}}\right). \tag{67}$$

Substituting these motion parameters and the intrinsic camera parameters into (56) and (57) yields

$$K_{\max} = \frac{6\ \text{mm}}{4.4\ \mu\text{m}} T_s \left\{ \frac{300\ \frac{\text{m}}{\text{s}}}{10\,000\ \text{m}} \right\} \tag{68}$$

$$K_{\max} = 40.9 T_s. \tag{69}$$

Applying the temporal sampling constraint and solving for T_s yields

$$T_s \leq \frac{1}{40.9} \quad (\text{s}) \tag{70}$$

which results in a minimum frame rate of 40.9 Hz and a maximum exposure time of 24.4 ms.

The final example represents the conditions expected during a low-level high-speed dash profile. As in the previous example, the camera is moving at 300 m/s with a relatively fixed orientation. However, in this case, the distance to the terrain is reduced to 300 m. The resulting motion parameters for this condition are given as follows:

$$\mathbf{v}^c = \begin{bmatrix} 300 \\ 0 \\ 0 \end{bmatrix} \left(\frac{\text{m}}{\text{s}}\right) \tag{71}$$

$$\boldsymbol{\omega}_{nc}^c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \left(\frac{\text{rad}}{\text{s}}\right). \tag{72}$$

Substituting these motion parameters and the intrinsic camera parameters into (56) and (57) yields

$$K_{\max} = \frac{6\ \text{mm}}{4.4\ \mu\text{m}} T_s \left\{ \frac{300\ \frac{\text{m}}{\text{s}}}{300\ \text{m}} \right\} \tag{73}$$

$$K_{\max} = 1363.6 T_s. \tag{74}$$

Applying the temporal sampling constraint and solving for T_s yields

$$T_s \leq \frac{1}{1363.6} \quad (\text{s}) \tag{75}$$

which results in a minimum frame rate of 1363.6 Hz and a maximum exposure time of 733 μs.

In this example, the relationship between the of velocity of the camera and the range to the target becomes apparent. To first order, the apparent motion of a feature in an image due to translational motion is proportional to the ratio of the velocity of the camera and the range to the target. Determining the minimum sampling constraints for general motion requires knowledge of the maximum velocity-to-range ratio expected based on the imaging conditions, as well as the maximum rotational rates between the camera and the world reference frame.

These case studies illustrate the frame rates required to sample at the Nyquist frequency. In general, the desired frame rates are not readily attainable using common hardware and lighting conditions. In many current correspondence search schemes (e.g., [10], [22], and [23]), the Nyquist sampling frequency for point sources is simply ignored, and the search scheme seeks the so-called "strong" features that are consistent between frames and geometrically consistent within a collection of other features [e.g., RANdom SAmple Consensus (RANSAC)]. It is our assertion that these feature extraction and correspondence techniques are effectively applying low-pass anti-temporal-aliasing filters that eliminate the higher frequency components that are corrupted by temporal aliasing.

As mentioned previously, there is a strong coupling between changes in camera pose and the apparent pixel motion. In the next section, measurements from an inertial sensor are used to mitigate the effects of temporal aliasing.

VIII. Incorporation of Inertial Sensor Measurements

As shown in the previous section, nonaliased temporal sampling can require relatively high frame rates, even for relatively simple imaging scenarios. High frame rates can present a number of challenges for a given imager, including high communication bandwidth requirements and short exposure times, requiring more sensitive (and expensive) sensors. We propose to exploit the information provided by inertial sensors to reduce the image sampling rates required to deliver antialiased measurements. The development of the aided sampling theory is presented as follows.

Inertial sensors can provide 3-D measurements of both the angular rate and the specific force (i.e., the sum of acceleration with respect to inertial and gravity) [24]. When combined with a kinematic model, this information can be exploited to produce an estimate of the trajectory. For the purposes of this illustration, the error dynamics can sufficiently be modeled using the following method.

When target motion was assumed to be effectively stationary, the apparent pixel motion [see (56) and (57)] was a function of the camera rotation rate and velocity with respect to the navigation frame and the relative location of the landmark. Strapdown inertial sensors measure both the angular rotation increment $\boldsymbol{\Delta\theta}_{ic}^c$ and the specific force increment $\boldsymbol{\Delta}\mathbf{v}^c$ with respect to the inertial reference frame. When combined with knowledge of the gravity vector, kinematic equations can be used to estimate the position, velocity, and attitude of the sensor. The inertial measurement errors, initial navigation state uncertainty, and errors in the gravity model all contribute to the inevitable unstable error growth experienced by all unaided strapdown INSs. A thorough development of these properties can be found in [24].

While all INSs experience unstable error growth over time, the relatively short durations between images allow us to model the errors between successive images using a simpler model. The first approximation assumes that the navigation reference frame is effectively an inertial reference frame over the short term. The second approximation assumes a general knowledge of the navigation state (e.g., the system is reasonably aligned) such that any errors in the navigation state itself do not dominate the pixel motion prediction between frames. Finally, in the cases where the camera is experiencing translational motion, prior knowledge of the range to the target is implicitly assumed. This information can be obtained using a number of techniques, both passive and active. See [25] for more details.

Thus, the simplified inertial sensor model represents the measurement as the sum of the true value plus an error and is given as

$$\tilde{\boldsymbol{\omega}}^c_{nc} = \boldsymbol{\omega}^c_{nc} + \boldsymbol{\delta\omega}^c_{nc} \tag{76}$$

$$\tilde{\mathbf{v}}^c = \mathbf{v}^c + \boldsymbol{\delta}\mathbf{v}^c \tag{77}$$

where $\boldsymbol{\omega}^c_{nc}$ is the true angular rotation rate, and $\mathbf{v}^c$ is the true velocity. The tilde represents the corrupted measurement as received from the inertial sensor. The inertial measurement errors $\boldsymbol{\delta\omega}^c_{nc}$ and $\boldsymbol{\delta}\mathbf{v}^c$ can be represented as random vectors with the following statistics over the interval T_s:

$$\mathbf{E}\left[\boldsymbol{\delta\omega}^c_{nc}\right] = \mathbf{0}_{3\times 3} \tag{78}$$

$$\mathbf{E}\left[\boldsymbol{\delta\omega}^c_{nc}\boldsymbol{\delta\omega}^{c^T}_{nc}\right] = q_w \tag{79}$$

$$\mathbf{E}[\boldsymbol{\delta}\mathbf{v}^c] = \mathbf{0}_{3\times 3} \tag{80}$$

$$\mathbf{E}\left[\boldsymbol{\delta}\mathbf{v}^c\boldsymbol{\delta}\mathbf{v}^{c^T}\right] = \left(\sigma^2_{v_0} + q_a T_s\right)\mathbf{I}_{3\times 3}. \tag{81}$$

The gyroscopic and accelerometer error sources are assumed to be collectively independent. Substituting the velocity and angle increment measurements from the inertial sensor algorithm into the pixel motion equations from (56) and (57) and integrating the error terms results in the residual pixel motion error rate due to inertial measurement errors, i.e.,

$$\delta\dot{u} = -\frac{f_0}{\Delta_x}\left(-\delta\omega^c_{nc_y} - \frac{\delta v^c_x}{s^c_z} + \frac{s^c_x s^c_y}{(s^c_z)^2}\delta\omega^c_{nc_x} - \left(\frac{s^c_x}{s^c_z}\right)^2 \times \delta\omega^c_{nc_y} + \frac{s^c_y}{s^c_z}\delta\omega^c_{nc_z} + \frac{s^c_x}{(s^c_z)^2}\delta v^c_z\right) \tag{82}$$

$$\delta\dot{v} = \frac{f_0}{\Delta_y}\left(\delta\omega^c_{nc_x} - \frac{\delta v^c_y}{s^c_z} - \frac{s^c_x s^c_y}{(s^c_z)^2}\delta\omega^c_{nc_y} + \left(\frac{s^c_y}{s^c_z}\right)^2 \times \delta\omega^c_{nc_x} - \frac{s^c_x}{s^c_z}\delta\omega^c_{nc_z} + \frac{s^c_y}{(s^c_z)^2}\delta v^c_z\right) \tag{83}$$

where $\delta\dot{u}$ and $\delta\dot{v}$ are the random pixel location errors rates in the x- and y-directions, respectively. The standard deviation of the residual pixel errors is given by calculating the variance of pixel errors after integrating over an interval of T_s, yielding

$$\sigma_u = \frac{f_0}{\Delta_x}\left[T_s\left(q_w + \frac{\sigma^2_{v_0} + q_a T_s}{s^c_z} + \frac{s^c_x s^c_y}{(s^c_z)^2}q_w + \left(\frac{s^c_x}{s^c_z}\right)^2 q_w + \frac{s^c_y}{s^c_z}q_w + \frac{s^c_x\left(\sigma^2_{v_0} + q_a T_s\right)}{(s^c_z)^2}\right)\right]^{1/2} \tag{84}$$

$$\sigma_v = \frac{f_0}{\Delta_y}\left[T_s\left(q_w + \frac{\sigma^2_{v_0} + q_a T_s}{s^c_z} + \frac{s^c_x s^c_y}{(s^c_z)^2}q_w + \left(\frac{s^c_y}{s^c_z}\right)^2 q_w + \frac{s^c_x}{s^c_z}q_w + \frac{s^c_y\left(\sigma^2_{v_0} + q_a T_s\right)}{(s^c_z)^2}\right)\right]^{1/2}. \tag{85}$$

The temporal sampling constraint can be applied in a similar manner as before; however, in this case, the constraint is applied to the standard deviation of *residual error* of pixel motion versus the total pixel motion considered in the unaided case, i.e.,

$$\sigma_{K_{\max}} = \sqrt{\sigma^2_u + \sigma^2_v}. \tag{86}$$

Enforcing the temporal sampling constraint on the residual random pixel motion requires selecting a confidence interval such that the residual pixel motion is constrained to less than one pixel uncertainty. This can be accomplished by evaluating the resulting probability distribution function of the residual pixel errors.

The preceding development is illustrated using a simple example. In this example, a consumer-grade inertial sensor is available with the following random walk parameters:

$$q_w = 4.2 \times 10^{-7}\ \frac{\mathrm{rad}^2}{\mathrm{s}} \tag{87}$$

$$q_a = 1.9 \times 10^{-5}\ \frac{\left(\frac{\mathrm{m}}{\mathrm{s}}\right)^2}{\mathrm{s}}. \tag{88}$$

As a further simplification, the pan components are isolated by assuming relatively distant targets (e.g., $s^c_z \to \infty$). This results in the following pixel uncertainties:

$$\sigma_u = \frac{f_0}{\Delta_x}[T_s q_w]^{1/2} \tag{89}$$

$$\sigma_v = \frac{f_0}{\Delta_y}[T_s q_w]^{1/2}. \tag{90}$$

Applying a 3-σ bound to the prediction errors results in the following temporal sampling constraint:

$$3\sigma_{K_{\max}} = 3f_0[T_s q_w]^{1/2}\sqrt{\left(\frac{1}{\Delta_x}\right)^2 + \left(\frac{1}{\Delta_y}\right)^2} \leq 1. \tag{91}$$

For simplicity, assume the pixel sizes are equivalent ($\Delta_x = \Delta_y = \Delta_{\text{pixel}}$), i.e.,

$$3\sigma_{K_{\max}} = \frac{3f_0[2T_s q_w]^{1/2}}{\Delta_{\text{pixel}}} \leq 1. \tag{92}$$

TABLE II
COMPARISON OF AIDED AND UNAIDED TEMPORAL IMAGE SAMPLING REQUIREMENTS FOR THE BASELINE CAMERA CONFIGURATION. SCENARIO DESCRIPTIONS ARE PROVIDED IN SECTION VII

Scenario Description	Unaided Sampling Rate (Hz)	Inertially Aided Sampling Rate (Hz)
Horizontal pan	≥ 146.7	≥ 14.06
High-altitude cruise	≥ 40.9	≥ 14.06
Low-altitude dash	≥ 1363.6	≥ 14.06

Solving for the sampling interval yields

$$T_s \leq \frac{\Delta^2_{\text{pixel}}}{18 f_0^2 q_w} \tag{93}$$

which corresponds to $\gamma = 36/\pi$ (derived in Section V). From Fig. 6, this sampling constraint captures approximately 98% of the signal energy, depending on the frequency response characteristics of the lens.

Substituting the previously presented camera and inertial parameters yields

$$T_s \leq \frac{1}{14.06} \text{ (s)} \tag{94}$$

which results in a minimum frame rate of 14.06 Hz and a maximum exposure time of 71.1 ms.

This illustration shows the benefits possible when utilizing inertial measurements to reduce temporal aliasing. This result implies that, as long as the rotational motion is within the bandwidth of the inertial sensor, sampling at ≥14.06 Hz will give acceptable antialiased results. Thus, when incorporating inertial measurements for feature anti-temporal aliasing, the sampling rate is effectively *independent* of camera motion.

With this result in mind, the required temporal sampling requirements can be compared between the unaided and inertially aided techniques. The results are shown in Table II. Tightly coupling inertial measurements can significantly reduce the image sampling frequency for nonaliased feature tracking and is effectively independent of the motion scenario. While this result is representative of a consumer-grade inertial sensor and camera, the required sampling rate is approximately inversely proportional to the quality of the inertial sensor. Thus, if a lower quality inertial sensor was used, q_w and q_a would increase, which as shown in (84) and (85), would result in higher required sample rates to eliminate temporal aliasing.

IX. CONCLUSION AND FUTURE WORK

In this paper, the concepts relating spatial and temporal image sampling have been explored from first principles with focus on the consequences for the correspondence search and feature tracking problem. The sampling theory has been developed and shown to yield a natural interrelationship between acceptable spatial and temporal sampling frequencies. As a result, we have shown that when point source features move more than one pixel between frames, spatial-temporal aliasing *is* occurring. The relationships between apparent feature motion and temporal sampling requirements have been shown to require very high (possibly unattainable) temporal sampling rates to guarantee nonaliased sampling. We believe that this is an underlying cause that forces designers to exploit complicated feature tracking algorithms, which, in essence, can be viewed as sophisticated anti-temporal-aliasing filters. Unfortunately, these popular feature tracking algorithms are *ad hoc* in the sense that they do not directly treat the temporal aliasing problem from a statistical sampling perspective.

Once the problem is posed from this perspective, the incorporation of inertial sensors is a natural choice. Inertial sensors have been shown to have the capability to statistically constrain the apparent motion effects, which can result in a significant reduction in the required temporal sampling rates while alleviating the burden of feature correspondence search. In essence, inertial sensors have been proposed as a direct method for reducing or eliminating temporal aliasing, allowing for the use of sophisticated and efficient/robust correspondence search algorithms and operation under lower lighting conditions.

Indeed, the use of inertial measurements for aiding the feature correspondence search task is akin to the use of inertial measurements in an ultratightly coupled Global Positioning System and INS, where the inertial information is used to steer the phase-locked loops in a feedforward mechanization. This facilitates precise code tracking under dynamic conditions—a powerful combination of precision and robustness that is the hallmark of properly fused synergistic sensors [26].

There are a number of issues which require further work and development. First, applying statistical constraints from inertial sensors requires some knowledge of the scene to properly account for translational motion. We propose to address this issue by incorporating statistical knowledge of the terrain (either *a priori* or *in situ*), which could be dynamically applied to either control temporal sampling rate or exclude features for which temporal aliasing is predicted.

Second, this development does not exploit any geometric constraints regarding the scene itself. In certain cases (e.g., an aircraft imaging a relatively flat scene), the temporal sampling rate can be reduced below the worst-case threshold presented in this paper.

Ultimately, we believe that this theory demonstrates the complimentary nature of imaging and inertial sensors. As such, properly incorporating inertial sensors can be a major advantage in developing robust image tracking applications within reasonable imaging and image processing constraints. Our goal is to use the theory developed in this paper to serve as a foundation for future research, where we will analyze current feature matching algorithms from a statistical sampling perspective and hopefully find ways to improve their performance.

REFERENCES

[1] M. J. Veth and M. Pachter, "Correspondence search mitigation using feature space anti-aliasing," in *Proc. Inst. Navigat. Annu. Meeting*, Apr. 2007, pp. 522–533.

[2] M. Pachter and A. Porter, "INS aiding by tracking an unknown ground object—Theory," in *Proc. Amer. Control Conf.*, 2003, vol. 2, pp. 1260–1265.

[3] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision*. New York: Springer-Verlag, 2004.

[4] S. Hrabar and G. S. Sukhatme, "A comparison of two camera configurations for optic-flow based navigation of a UAV through urban canyons," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2004, vol. 3, pp. 2673–2680.

[5] C. F. Olson, L. H. Matthies, M. Schoppers, and M. W. Maimone, "Robust stereo ego-motion for long distance navigation," in *Proc. IEEE Conf. Adv. Robot.*, Jun. 2000, vol. 2, pp. 453–458.

[6] H. Adams, S. Singh, and D. Strelow, "An empirical comparison of methods for image-based motion estimation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2002, vol. 1, pp. 123–128.

[7] M. Pachter and A. Porter, "Bearings-only measurements for INS aiding: The three-dimensional case," presented at the AIAA Guidance, Navigation Control Conf., Austin, TX, 2003, AIAA Paper 2003-5354.

[8] E. Hagen and E. Heyerdahl, "Navigation by optical flow," in *Proc. 11th IAPR Int. Conf. Pattern Recog.*, 1992, vol. 1, pp. 700–703.

[9] J. F. Raquet and M. Giebner, "Navigation using optical measurements of objects at unknown locations," in *Proc. 59th Annu. Meeting Inst. Navigat.*, Jun. 2003, pp. 282–290.

[10] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. DARPA Image Understanding Workshop*, 1981, pp. 121–130.

[11] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. Comput. Vis.*, Corfu, Greece, Sep. 1999, vol. 2, pp. 1150–1157.

[12] B. Bhanu, B. Roberts, and J. Ming, "Inertial navigation sensor integrated motion analysis for obstacle detection," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 1990, pp. 954–959.

[13] D. W. Strelow, "Motion estimation from image and inertial measurements," Ph.D. dissertation, Sch. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, Nov., 2004.

[14] M. J. Veth, J. F. Raquet, and M. Pachter, "Stochastic constraints for efficient image correspondence search," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 973–982, Jul. 2006.

[15] M. J. Veth and J. F. Raquet, "Two-dimensional stochastic projections for tight integration of optical and inertial sensors for navigation," in *Proc. Inst. Navigat. Nat. Tech. Meeting*, 2006, pp. 587–596.

[16] J. W. Goodman, *Introduction to Fourier Optics*. Boston, MA: McGraw-Hill, 1996.

[17] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[18] E. Hecht, *Optics*. San Fransisco, CA: Addison-Wesley, 2002.

[19] R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial Value Problems*. Malabar, FL: Krieger, 1994.

[20] M. J. Veth and J. F. Raquet, "Alignment and calibration of optical and inertial sensors using stellar observations," in *Proc. ION GNSS*, Sep. 2005, pp. 2494–2503.

[21] A. M. Tekalp, *Digital Video Processing*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[23] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. CVPR*, 2004, vol. 2, pp. 506–513.

[24] D. Titterton and J. Weston, *Strapdown Inertial Navigation Technology*. Lavenham, U.K.: Peregrinus, 1997.

[25] J. Morrison, J. F. Raquet, and M. J. Veth, "Vision aided inertial navigation system augmented with a coded aperture," in *Proc. ION Nat. Tech. Meeting*, 2009, pp. 61–73.

[26] M. E. Oxley and S. N. Thorsen, "Fusion or integration: What's the difference?," in *Proc. 7th Int. Conf. Inf. Fusion*, Jun. 2004, pp. 429–434.

Michael J. Veth (SM'09) received the B.S. degree in electrical engineering from Purdue University, West Lafayette, IN, and the Ph.D. degree in electrical engineering from the Air Force Institute of Technology (AFIT), Dayton, OH. In addition, he is a graduate of the Air Force Test Pilot School, Dayton.

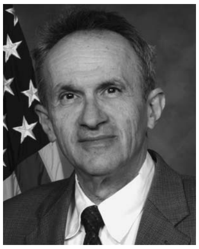
He is currently an Assistant Professor of electrical engineering with the Department of Electrical and Computer Engineering, AFIT, where he serves as the Deputy Director of the Advanced Navigation Technology Center. His current research focus is understanding and implementing bio-inspired methods to fuse image and inertial systems for navigation, targeting, and control.

Dr. Veth is a member of the Institute of Navigation, Tau Beta Pi, and Eta Kappa Nu.

Richard K. Martin (M'09) received dual B.S. degrees (*summa cum laude*) in physics and electrical engineering from the University of Maryland, College Park, in 1999 and the M.S. and Ph.D. degrees in electrical engineering from Cornell University, Ithaca, NY, in 2001 and 2004, respectively.

Since August 2004, he has been with the Department of Electrical and Computer Engineering, Air Force Institute of Technology (AFIT), Dayton, OH, where he is an Associate Professor. He is the author of 20 journal papers and 40 conference papers. He is the holder of four patents. His research interests include navigation and source localization; cognitive radio; equalization for cyclic-prefixed systems; blind, adaptive filters; sparse adaptive filters; and laser radar.

Dr. Martin has been elected Electrical and Computer Engineering Instructor of the Quarter three times and Eta Kappa Nu Instructor of the Year twice by the AFIT students.



Meir Pachter (F'00) received the B.S. and M.S. degrees in aerospace engineering and the Ph.D. degree in applied mathematics from the Israel Institute of Technology, Haifa, Israel, in 1967, 1969, and 1975, respectively.

He is currently a Professor of electrical engineering with the Department of Electrical and Computer Engineering, Air Force Institute of Technology, Dayton, OH. He has held research and teaching positions with the Israel Institute of Technology; the Council for Scientific and Industrial Research in South Africa; the Virginia Polytechnic Institute, Blacksburg; Harvard University, Cambridge, MA; and Integrated Systems, Inc. His current areas of interest include statistical signal processing, adaptive optics, inertial navigation, and Global Positioning System navigation.

Dr. Pachter received the Air Force Air Vehicle Directorate Foulois award in 1994 for his work on adaptive and reconfigurable flight control.