



ELSEVIER

Infrared Physics & Technology 38 (1997) 177–192

INFRARED PHYSICS
& TECHNOLOGY

Restoration and range performance in FLIR imaging systems

Stephen E. Reichenbach^{a,*}, Stephen K. Park^b, James D. Howe^c

^a *Computer Science and Engineering Department, University of Nebraska-Lincoln, Lincoln, NE 68588-0115, USA*

^b *Computer Science Department, College of William and Mary, Williamsburg, VA 23185, USA*

^c *Night Vision and Electronic Sensors Directorate, US Army, FT. Belvoir, VA 22060, USA*

Received 3 May 1996

Abstract

Image restoration of forward-looking infrared (FLIR) imagery has the potential to significantly improve the quality of images used by an automatic target recognizer (ATR) or human observer. This study investigates the feasibility of real-time image restoration algorithms and the problem of measuring image quality as it relates to target acquisition performance. This paper describes a technique for deriving small kernel filters that efficiently restore and reconstruct. Subject to implementation constraints associated with efficient application, the filters optimize image fidelity to an 'ideal' close-range image. The paper describes simulation experiments employing an end-to-end imaging system model, experiments with actual images using a model-based characterization of an actual imaging system, and simulation experiments that illustrate the utility of the system model and filtering in FLIR imaging system design. © 1997 Published by Elsevier Science B.V.

1. Introduction

Performance prediction is an important task in the design of forward-looking infrared (FLIR) imagers [1]. In FLIR systems, performance usually refers to target acquisition, including detection, classification, recognition, and identification either by a human observer or automatic target recognizer (ATR). The distance at which a human observer or ATR can reliably identify an object in the image is a fundamental measure of performance. Digital image processing offers the potential to improve the effective range in such systems by improving the quality of the digital image presented to the observer or ATR. This paper presents an efficient digital filter for enhancing image quality to improve range performance and investigates FLIR system design tradeoffs

that affect image quality and range performance.

Image acquisition inevitably introduces degradations [2]. The goal of digital image restoration is to recover the 'true' image by removing these degradations [3]. It is widely recognized that FLIR images (and other remotely sensed images) are blurred by the atmosphere, optics, and detectors. Early methods for digital image restoration were designed to correct such blurring. Researchers and practitioners soon realized that attempts to invert the blurring process must account for image noise, including shot noise, detector noise, and circuit noise [3]. More recent research has demonstrated that sampling and reconstruction also have great significance for image quality in digital systems and should be addressed by restoration methods [4,5].

Sampling, the process of obtaining irradiance measures at spatially discrete points, is fundamental to range performance. The number of samples on target (i.e., the target size measured in pixels) is one

* Corresponding author. E-mail: reich@cse.unl.edu.

of the most critical issues in recognition. For a specific imaging system configuration, the number of samples on target in each dimension is inversely proportional to range (i.e., if the range is doubled, the number of pixels on target in each dimension is halved). For a given field-of-view (FOV), sampling density is necessarily limited by available technology and cost. Systems with insufficient sampling density introduce aliasing artifacts. For a fixed sampling density, the only way to reduce aliasing is with pre-sample blurring (e.g., using the optics). Of course, pre-sample blurring typically also degrades the image by attenuating components within the Nyquist limit.

Although the tradeoff between pre-sample blurring and aliasing is a fundamental issue that has a large effect on the overall performance of digital imaging systems, it is generally ignored in the literature [2]. In 1976, Ratches identified the failure to account for aliasing and sampling effects as a critical shortcoming in performance models of thermal imaging systems [6]. A decade later, McCracken and Wajsfelner identified the same problem [7]. Recently, Kennedy reiterated the need to account for sampling and aliasing in performance models of second-generation thermal imaging systems [8].

This paper describes the derivation of an efficient filter that not only restores (i.e., corrects for systemic degradations), but also interpolates between pixel values to improve reconstruction and thereby increase apparent sharpness and clarity. The approach is based on a comprehensive imaging system model that accounts for the statistics of the scene, the image-acquisition blurring, sampling, noise, and post-filter reconstruction [4,5]. The derivation is conditioned on explicit constraints on spatial support and filter density, insuring efficient implementation by convolution [9]. Subject to these constraints, the resulting filter optimizes the fidelity of the output

image to an ‘ideal’ close-range image. The paper investigates system design tradeoffs that affect filter effectiveness and overall system performance.

2. Digital imaging system model and notation

Restoration requires modeling the system. This section presents an end-to-end model of the digital imaging process, illustrated in Fig. 1. The model accounts for the transformation from continuous to discrete data during image acquisition and from discrete to continuous data during image reconstruction. This continuous–discrete–continuous (CDC) model is a more accurate model for digital imaging systems than are more conventional discrete–discrete or continuous–continuous models and is the basis for the development of significantly more effective filters.

The end-to-end digital imaging process consists of three phases:

- image acquisition, in which the digital image acquisition device (e.g., a FLIR camera) converts the continuous radiance field s associated with a scene into a digital image p that is both sampled and quantized;
- digital processing, in which a digital filter is applied to the digital image p to produce an improved digital image q ; and
- post-filter reconstruction, where further analog and/or digital processing yields a spatially continuous representation r .

The filter f described in this paper is implemented by discrete convolution, so both the input and output of the filter (p and q respectively) are digital images. The filter restores the image by correcting for systemic degradations and also interpolates between pixel values to improve reconstruction. In a typical system, the visual effect is increased sharpness and clarity. So, for example, for a filter defined on a

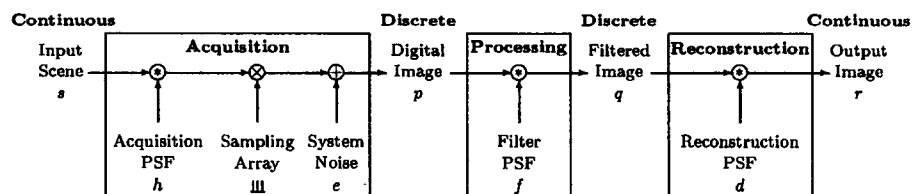


Fig. 1. Mathematical model of the digital imaging process.

lattice with four times the sampling density, a 64×64 input image would yield a 256×256 output image with increased element density.

Depending on the application, post-filter reconstruction may be implemented either (1) with an analog/optical device such as a display monitor for human viewing and interpretation; (2) with post-filter digital processing (e.g., bilinear interpolation) for geometric correction, translation, scaling, rotation, remapping, or other geometric operations requiring resampling; or (3) not at all. Physical display devices inherently reconstruct; some ATRs implement digital reconstruction (e.g., resizing the image for a bank of matched filters) and some do not.

The digital image p is

$$p[n] = \int_{-\infty}^{\infty} h(n-x)s(x)dx + e[n] \quad (1)$$

where the continuous scene radiance field s is convolved with the pre-sampling acquisition point spread function (PSF) h (e.g., characterizing blurring from the optics and the spatial integration of the photodetector) and degraded by noise e (e.g., shot noise, circuit noise, and quantization error). In reality, noise is caused by spatially continuous processes, but one can define a discrete noise process that has statistically identical effects on the digital image. This and subsequent equations are written in one dimension for notational convenience, but generalize directly to multiple dimensions. Pixels are indexed with integer coordinates $[n]$ and the continuous spatial coordinates (x) are normalized to the sampling interval (i.e., the intersample distance is 1). Function values are expressed on a gray-level equivalent scale.

In the spatial-frequency domain,

$$\hat{p}(v) = \sum_{k=-\infty}^{\infty} \hat{h}(v-k)\hat{s}(v-k) + \hat{e}(v) \quad (2)$$

where spatial frequencies v are normalized to the sampling frequency, \hat{p} is the Fourier transform or spatial-frequency spectrum of the image, \hat{h} is the acquisition transfer function, \hat{s} is the spatial-frequency spectrum of the scene, and \hat{e} is the spatial-frequency spectrum of the noise. Sampling causes the folding of the components of the spatial-frequency spectrum. The resulting image transform \hat{p} is periodic with period equal to the normalized sampling frequency, 1.0 in each dimension. (In the digital

image, the noise is discrete, so its spectrum \hat{e} is also periodic.)

Filtering produces the filtered digital image q from the input digital image p . If the digital filter has the same density as the input image, then the filtered image will have the same pixel density as the unprocessed digital image. However, the digital filter need not have the same density as the input image. This allows for reconstructing images with greater density than the input image during restoration. For example, consider the one-dimensional image

$$p = [\dots 111555 \dots]$$

zero filled to twice the density

$$p = [\dots 10101050505 \dots]$$

then convolved with a filter with twice the density of the original image

$$f = \frac{1}{4}[\dots 00012221000 \dots].$$

The result is a filtered image that has twice the density of the input image

$$q = [\dots 11112345555 \dots]$$

For simplicity, we will consider only filters with element density that is an integer multiple of the density of the image p .

If we define the restoration and reconstruction filter f on a lattice that has R elements per pixel (indicating the coordinates of filter elements with integer indices divided by filter density), then the filtered image q is

$$q\left[\frac{n}{R}\right] = \sum_{n'=-\infty}^{\infty} f\left[\frac{n}{R} - n'\right]p[n']. \quad (3)$$

One can view this operation as convolution of the filter with an image that has zero values at points between pixels. As a practical matter, only a finite image p is available for processing and the spatial support of the digital filter f must be limited. To address these issues and to facilitate the use of the discrete Fourier transform (DFT), it is common to assume the scene and hence the image are periodic with period equal to the image size (i.e., the field of view) and to constrain the filter support to the size of the image (or smaller). This means the filtering operation is one of circular convolution which can be implemented using the DFT.

The frequency-domain equation for filtering (corresponding to Eq. (3)) is

$$\hat{q}(v) = \hat{f}(v) \hat{p}(v). \quad (4)$$

With the filter density limited to R elements per pixel, the filter transfer function \hat{f} is periodic with period R times the sampling frequency. (Just as spatial convolution can be envisioned with zero values filled between image pixels, so can frequency domain multiplication be envisioned with the periodic replication of the image transform \hat{p} to produce an array of the same length as the filter transform \hat{f} .) With the filter density a multiple of the sampling density, the pixel density of the filtered image q is the same as that of the filter, and its transform \hat{q} is also periodic with period R times the sampling frequency. With a periodic image and discrete filter, the transforms in Eq. (4) are discrete as well as periodic, and can be computed with the DFT.

Reconstruction produces a continuous image from a digital image. Even if the filter f interpolates to perform some reconstruction, subsequent components of the system such as a display device may implement post-filter reconstruction. For example, a display monitor produces a spot for each pixel value, effectively blurring the discrete pixel values to form a continuous image. The display spot is designed to produce uniformity in homogeneous regions without excessive blurring of sharp transitions [2]. Convolution with a Gaussian spot is commonly used to model the display operation of video monitors. Traditional digital resampling operations also implicitly convolve the digital image with a continuous reconstruction function even though the result is computed only at the resampled points. Nearest-neighbor, bilinear, and cubic convolution [10] are common digital reconstruction methods that use continuous reconstruction PSFs.

After convolution with the reconstruction PSF d , the continuous output image r is

$$r(x) = \sum_{n=-\infty}^{\infty} d\left(x - \frac{n}{R}\right) q\left[\frac{n}{R}\right]. \quad (5)$$

The corresponding frequency-domain equation for reconstruction is

$$\hat{r}(v) = \hat{d}(v) \hat{q}(v) \quad (6)$$

where \hat{d} is the reconstruction transfer function.

3. System performance

This section addresses the problem of defining a measure of system performance. Ideally, filters would be evaluated by determining the performance of the human or ATR given images passed through the filter. Unfortunately, the relationship between image imperfections and target recognition success is complex and not well understood [1]. Although many models of the process have been developed, most are specialized and are only partially validated by realistic field tests [11–14]. Although current prediction methods are bound to evolve and improve, a definitive theoretical method for evaluating filter performance does not seem possible at this time. Moreover, because there is no comprehensive or standard suite of ATRs or of test images for empirically assessing the effectiveness of filters in improving performance, a definitive empirical method also is not possible.

An alternative measure for filter effectiveness is how well it recovers the ‘ideal’ image – the scene radiance field. The perfect filter would not only restore sample values (i.e., remove degradations); it would also reconstruct the spatially continuous ‘ideal’ image of the scene from the spatially discrete pixel values. Such a perfect filter would provide an ideal projection of the scene to the human observer or ATR. To evaluate a filter, one ideally would compare the filter output to the scene radiance field. Unfortunately, such ideal images without degradation and with unlimited resolution cannot be acquired, represented, or manipulated in a digital computer and so this direct comparison is not possible. However, as described below, we can use nearly ideal images that, for our purposes, effectively are free of degradations and effectively have unlimited resolution.

Range is a principal determinant of performance – objects are more readily detected, recognized, and identified at close range. (We are primarily concerned here with static performance. Performance in the field is also affected by time, field-of-view, and other factors.) With most target acquisition systems, an image of an object taken at close range allows the best possible performance. Close-range images also are nearly ‘ideal’ images because the degradations such as blurring and noise are minor relative to the

level of detail and strength of signal in the image and because the sampling density (relative to the target) is great. Therefore, in experiments with both simulated and actual images, we use close-range images as the ideal images of the scenes. In order to be effectively ideal, the ratio between the sample density of the close-range image and the sample density of the test image should be great enough so that the close-range images are effectively continuous with respect to the imaging process. Also, the blurring in the close-range image should be small relative to the details in the scene and the signal-to-noise ratio (SNR) in the close-range image should be large.

With reference to the imaging model presented in Section 2, system performance will be measured by how closely the output image r matches the ideal (or close-range) image s . Linfoot [15] used the expected mean-square error of an imaging system (with stochastic scene and noise)

$$\begin{aligned} S^2 &= E \left\{ \int_{-\infty}^{\infty} |s(x) - r(x)|^2 dx \right\} \\ &= E \left\{ \int_{-\infty}^{\infty} |\hat{s}(v) - \hat{r}(v)|^2 dv \right\} \end{aligned} \quad (7)$$

to define image fidelity as

$$F = 1 - \frac{S^2}{\sigma_s^2} \quad (8)$$

where σ_s^2 is the expected (ensemble average) variance of the scene radiance field

$$\sigma_s^2 = E \left\{ \int_{-\infty}^{\infty} |s(x)|^2 dx \right\} = E \left\{ \int_{-\infty}^{\infty} |\hat{s}(v)|^2 dv \right\}. \quad (9)$$

For notational simplicity, equations in this paper assume the scene radiance field is a zero-mean process; in practice, the mean can be accounted for during filtering. Fidelity is bounded by 1, with $F = 1$ if and only if the output image is identical to the scene radiance field. Mean-square-error metrics such as fidelity are intuitive and facilitate mathematical analyses, but do not correspond directly to human assessments of image quality and are not related directly to performance by human observers or all ATRs. However, a more definitive objective measure of image quality has proven elusive.

Linfoot [15] defined two related measures of image quality: relative structural content

$$T = \sigma_r^2 / \sigma_s^2 \quad (10)$$

and correlation quality

$$Q = \sigma_{s,r} / \sigma_s^2 \quad (11)$$

where σ_r^2 is the expected variance of the output image and $\sigma_{s,r}$ is the expected covariance of the scene radiance field and image. Fidelity, relative structural content, and correlation quality are related by the equation

$$Q = \frac{1}{2}(F + T). \quad (12)$$

This is a corollary of the equation

$$S^2 = \sigma_s^2 - 2\sigma_{s,r} + \sigma_r^2. \quad (13)$$

The correlation coefficient is a related measure of the similarity between the scene and image:

$$C = \frac{\sigma_{s,r}}{\sigma_s \sigma_r} = \frac{Q}{\sqrt{T}}. \quad (14)$$

The correlation coefficient is invariant with respect to changes in offset and gain (additive and multiplicative scalar constants) and is used in template matching, one of the most fundamental object detection algorithms [16].

4. Digital filter design

This section presents the derivation of an efficient, high-density digital filter for improved imaging system performance. The filter improves upon more traditional filters because it is based on the comprehensive, end-to-end model described in Section 2. The filter improves image quality in two ways:

(1) by restoration, addressing acquisition artifacts including presample blur, aliasing, and noise, and

(2) by partial reconstruction, increasing apparent sharpness and clarity by digital interpolation.

The filter is efficient because spatial support and density are constrained during the design process. Subject to these constraints, the derivation yields the filter that optimizes fidelity for the end-to-end system model.

The two major considerations in the design of the

digital filter f are effectiveness and efficiency. As one would expect, there is a tradeoff between effectiveness and efficiency: achieving efficiency may involve some loss of effectiveness. A common approach to balancing these two competing demands is to design a filter for optimal effectiveness and then modify it ad hoc to satisfy efficiency constraints (e.g., by truncating or windowing the spatial filter). However, that approach sacrifices the rationale used in defining the filter. Here, we impose explicit support and density constraints in the derivation of the filter.

If the digital restoration filter is defined with greater density than the digital image (i.e., the distance between filter elements is less than the pixel interval), then the filtering operation interpolates as well as restores the image. For a digital image with N pixels, a filter with density R elements per pixel (in each dimension) produces a filtered image with RN pixels (in each dimension). If the subsequent reconstruction function d is defined relative to the filter density, then as the filter density R increases, the relative effect of subsequent reconstruction with d is diminished. In the limit, the filter f is continuous and the post-filter reconstruction function d is the impulse or delta function (and has no effect). Although increasing the filter density improves system performance by improving reconstruction, the cost is increased computation.

The computation required to apply a filter is a function of the size of its spatial support and its density. (The spatial support defines the number and location of the filter elements.) The more elements in the filter support, the more computation is required for spatial convolution. The computational complexity for convolution is $O(KN)$ where K is the number of elements in the filter and N is the number of image pixels. (In this analysis, N is the total number of pixels in the image regardless of the image dimensionality.) Similarly, the spatial density of the filter determines the computational complexity of Fourier domain processing. The computational complexity for the Fourier transform is $O(RN \log(RN))$ where R is the density of the filter.

If the filter f is constrained to a kernel of a few elements, it can be applied very efficiently by convolution. Small kernel filters are also well suited for parallel, focal-plane implementation [9].

The design optimization criterion is to minimize mean-square error (or equivalently to maximize fidelity) of the system. The derivation assumes that noise is signal independent and that sidebands of the scene spectrum that alias to the same frequency are uncorrelated. Then, the terms of the expression for expected mean-square error in Eq. (13) can be expressed as

$$\sigma_s^2 = \int_{-\infty}^{\infty} \hat{\Phi}_s(v) dv, \quad (15)$$

$$\sigma_{s,r} = \int_{-\infty}^{\infty} \hat{\Phi}_{s,p}(v) \hat{f}^*(v) \hat{d}^*(v) dv, \quad (16)$$

$$\sigma_r^2 = \int_{-\infty}^{\infty} \hat{\Phi}_p(v) |\hat{f}(v)|^2 |\hat{d}(v)|^2 dv \quad (17)$$

where the asterisk superscript denotes complex conjugation and

$$\hat{\Phi}_s(v) = E\{|\hat{s}(v)|^2\}, \quad (18)$$

$$\hat{\Phi}_{s,p}(v) = E\{\hat{s}(v) \hat{p}^*(v)\} = \hat{\Phi}_s(v) \hat{h}^*(v), \quad (19)$$

$$\begin{aligned} \hat{\Phi}_p(v) &= E\{|\hat{p}(v)|^2\} \\ &= \sum_{k=-\infty}^{\infty} \hat{\Phi}_s(v - kR) |\hat{h}(v - kR)|^2 + \hat{\Phi}_e(v). \end{aligned} \quad (20)$$

As stated, in actual practice using finite images on a digital computer, we assume scene periodicity so the continuous integrals in Eqs. (15)–(17) are replaced with summations. Because all spatial functions are real, the frequency-domain transforms are Hermitian and the power spectra $\hat{\Phi}_s$, $\hat{\Phi}_p$, and $\hat{\Phi}_e$ are real, non-negative, and even. The image and noise spectra are periodic with the sampling frequency.

With no restrictions on the density or spatial size of the filter, the mean-square error is minimized when the filter transfer function is [4]

$$\hat{f}_w(v) = \frac{\hat{\Phi}_{s,p}(v) \hat{d}^*(v)}{\hat{\Phi}_p(v) |\hat{d}(v)|^2}. \quad (21)$$

We refer to this filter as the CDC Wiener filter because it is based on the continuous–discrete–continuous (CDC) model. This filter is not practical, however, because its support is the size of the image and it has unconstrained density (i.e., it is continuous). To produce a practical filter, we impose con-

straints on the density and spatial size prior to the derivation.

If the filter is limited to R elements/pixel, the optimal filter is [5]

$$\hat{f}_1(v) = \frac{\hat{b}(v)}{\hat{a}(v)} \quad (22)$$

where

$$\hat{b}(v) = \sum_{k=-\infty}^{\infty} \hat{\Phi}_{s,p}(v - kR) \hat{d}^*(v - kR), \quad (23)$$

$$\hat{a}(v) = \sum_{k=-\infty}^{\infty} \hat{\Phi}_p(v - kR) |\hat{d}(v - kR)|^2. \quad (24)$$

This filter transfer function is periodic with period R cycles/pixel. We refer to this filter as the unconstrained Wiener filter because it is spatially unconstrained.

We also restrict the spatial support of the filter to a small set of locations C on the lattice defined by the limiting density of the filter. (The number of locations in the constraint set is K .) The elements not in the support set C cannot be used in the filter; i.e., the filter is zero at these locations:

$$f_c\left[\frac{n}{R}\right] = 0 \quad \text{if} \quad \left[\frac{n}{R}\right] \notin C. \quad (25)$$

The integer R in the denominator of the location allows the filter to have greater density than the image – R elements per pixel in each dimension.

With density and support restricted, the optimal constrained filter is defined by

$$\sum_{[n'/R] \in C} a\left[\frac{n-n'}{R}\right] f_c\left[\frac{n'}{R}\right] = b\left[\frac{n}{R}\right] \quad (26)$$

for each $n \in C$ where a and b are the inverse transforms of \hat{a} and \hat{b} respectively. The number of unknowns in Eq. (26) is equal to the number of locations in the support set of the filter. If the constraint set C has K elements, there are K equations in K unknowns.

After applying this filter, the digital image will have increased density, but further reconstruction is required to produce continuous output. However, because the constrained filter will have accomplished the most significant reconstruction, the relative effects of any subsequent reconstruction should be largely diminished. For object recognition, the filter

and reconstruction operations might be applied only to areas of interest and not to the whole image to further reduce computation.

Eq. (13) with the expressions in Eqs. (15)–(17) can be used to compute the expected mean-square error from $\hat{\Phi}_s$, $\hat{\Phi}_e$, \hat{h} , \hat{f} , and \hat{d} for various systems. The expected mean-square error for a system with the unrestricted filter f_w in Eq. (21) reduces to

$$S_w^2 = \int_{-\infty}^{\infty} \hat{\Phi}_s[v] - \frac{\hat{\Phi}_{s,p}[v]}{\hat{\Phi}_p[v]} dv. \quad (27)$$

Therefore, the fidelity (Eq. (8)) for this filter is

$$F_w = \frac{1}{\sigma_s^2} \int_{-\infty}^{\infty} \frac{|\hat{\Phi}_{s,p}(v)|^2}{\hat{\Phi}_p(v)} dv. \quad (28)$$

No filter can restore with higher fidelity (smaller mean square error) than F_w in Eq. (28). However, for typical imaging systems, the optimal filter PSF has a few centrally located elements that account for most of the filter response and the most significant response is at lower frequencies, so it is reasonable to expect that small kernel filters with limited density can perform nearly as well. For any filter f , the expected mean square error is

$$S^2 = S_w^2 + \int_{-\infty}^{\infty} \hat{\Phi}_p(v) |d(v)|^2 |\hat{f}(v) - \hat{f}_w(v)|^2 dv. \quad (29)$$

As can be seen in Eq. (29) $S^2 \geq S_w^2$ and F_w is the upper bound on fidelity.

Other image quality metrics, including expected relative structural content, correlation quality, and correlation coefficient, can be computed in a similar fashion. In Section 6, we use these measures to evaluate various FLIR imaging system designs.

5. Modeling FLIR systems

In this section, we develop a simple linear, shift-invariant model for FLIR imaging systems based on the continuous/discrete/continuous model in Fig. 1. Despite the fact that this model does not account for all aspects of the imaging process, including shift-variant and non-linear characteristics of FLIR systems, it does provide the basis for deriving effective

filters and for experiments that yield substantial insight into overall system performance.

In previous sections of this paper, we used one-dimensional notation for simplicity. Here, we examine a two-dimensional system, so we introduce some additional notation. As before, all the system coordinates are normalized to the sampling interval Δ in the image plane in each dimension. For simplicity, we assume the horizontal and vertical sampling intervals are equal and the numbers of pixels in each dimension are equal. Then, object-plane coordinates (x_o, y_o) are mapped to normalized image-plane coordinates (x, y) as

$$x = \frac{D_i x_o}{D_o \Delta}, \quad y = \frac{D_i y_o}{D_o \Delta} \quad (30)$$

where D_i is the focal length and D_o is the distance to the object.

We model the autocorrelation of the scene as

$$\Phi_s(x, y) = \exp\left(-\sqrt{x^2 + y^2} / \bar{x}\right). \quad (31)$$

where the parameter \bar{x} is the mean spatial detail of the object or scene (in normalized dimension units). This function is the non-separable autocorrelation of a circularly symmetric Markov process and is widely used for modeling scene autocorrelation [17,18]. The corresponding power spectrum is

$$\hat{\Phi}_s(u, v) = \frac{2\pi \bar{x}^2}{(1 + 4\pi^2 \bar{x}^2 \omega^2)^{3/2}} \quad (32)$$

where (u, v) is the continuous normalized spatial frequency and $\omega = \sqrt{u^2 + v^2}$ is the radial frequency magnitude.

We model three components that contribute to the image acquisition PSF and transfer function: the optics, the detector, and the scanning mechanism. The model for the blur-degraded optical transfer function (OTF) is

$$\hat{h}_o(u, v) = \Pi(\omega/2\omega_0) \left(\cos^{-1}(\omega/\omega_0) - \frac{\omega}{\omega_0} \right. \\ \left. \times \sqrt{1 - (\omega/\omega_0)^2} \right) \exp\left(-\frac{1}{2}(\omega/\omega_b)^2\right) \quad (33)$$

where the OTF cutoff ω_0 for a circular exit pupil is

defined by the focal length D_i , aperture diameter A , mean wavelength λ , and the sampling interval Δ as

$$\omega_0 = \frac{A\Delta}{\lambda D_i}$$

with

$$\Pi(\omega) = \begin{cases} 1 & \omega < \frac{1}{2} \\ 0 & \omega \geq \frac{1}{2} \end{cases}$$

and optical defocus blur is modeled as a Gaussian spot with width ω_b (in the frequency domain). Multiplicative scale factors are collected in defining signal-to-noise ratio, given below. Other, more complex models for optical defocus can be found in [19,20].

We use a simple square pulse to model the detector PSF (to account for spatial integration by the detector). The size of the detector relative to the sampling array is given by the samples-per-dwell P . The transfer function of the detector is

$$\hat{h}_d(u, v) = \text{sinc}(Pu)\text{sinc}(Pv) \quad (34)$$

where

$$\text{sinc}(u) = \begin{cases} \frac{\sin(\pi u)}{\pi u} & u \neq 0 \\ 1 & u = 0 \end{cases}$$

Properly, the function in Eq. (34) should have a multiplicative scale factor of P^2 , but we collect all scale factors in defining signal-to-noise ratio (given below).

In a scanned system, the linear motion of the scanner causes a uniform-pulse PSF in the scan direction. With 100% duty-cycle, the scanner transfer function, in terms of normalized frequencies, is

$$\hat{h}_s(u, v) = \text{sinc}(u) \quad (35)$$

Systems with two-dimensional detector arrays may still have a one-dimensional blur associated with readout (e.g., as in [21]).

The composite acquisition transfer function of a scanning system is

$$\hat{h}(u, v) = \hat{h}_o(u, v)\hat{h}_d(u, v)\hat{h}_s(u, v). \quad (36)$$

In a staring system, there is no scanning motion and so \hat{h}_s is not included.

We use a simple additive white noise model for system noise (i.e., the noise power spectrum $\hat{\Phi}_e$ is constant across all spatial frequencies in the digital

image). A number of factors affect the signal-to-noise ratio (SNR), including the strength of the scene radiance field, atmospheric attenuation, sampling density, detector size and responsivity variations, and various sources of system noise [22–24]. SNR is defined as the ratio of the root-mean-square (RMS) scene signal σ_s to the RMS noise σ_e

$$\text{SNR} = \sigma_s / \sigma_e. \quad (37)$$

The SNR can be varied in the model to account for factors such as atmospheric attenuation of the signal, sensor detectivity, detector size, gain, and electrical bandwidth.

The simulations presented in Section 6 employ four reconstruction functions: a Gaussian spot to model the display spot of a display device, nearest neighbor, bilinear interpolation, and cubic convolution [10]. The transfer function of a Gaussian spot with a spread half the distance between elements in the filtered image q is

$$\hat{d}(u, v) = \exp\left(-\frac{1}{2}(R\pi w)^2\right). \quad (38)$$

Alternatively, the transfer function of nearest neighbor interpolation (multidimensional generalization of the sample-and-hold square pulse) is

$$\hat{d}(u, v) = \text{sinc}(Ru)\text{sinc}(Rv), \quad (39)$$

the transfer function of the bilinear interpolation function is

$$\hat{d}(u, v) = \text{sinc}^2(Ru)\text{sinc}^2(Rv), \quad (40)$$

and the transfer function of the separable cubic convolution reconstruction function is

$$\hat{d}(u, v) = \hat{d}_p(Ru)\hat{d}_p(Rv) \quad (41)$$

where

$$\hat{d}_p(v) = (\pi v)^{-2}(3 \text{sinc}^2(v) - \text{sinc}(2v) - 3 \text{sinc}^2(2v) + \text{sinc}(4v)). \quad (42)$$

This parametric model of FLIR systems is the basis for experiments described in Section 6 that compare filter performance for various FLIR system designs. As has already been noted, the model does not account for all aspects of FLIR systems. Nonetheless, we believe that it captures the most significant factors affecting system performance and therefore provides the basis for deriving effective filters and conducting valid experiments.

6. Experimental results

This section presents three sets of experimental results that assess filter effectiveness and overall system performance under a range of conditions. Results with both simulated images and actual images are presented. The software simulation allows precise analyses of restoration and reconstruction performance; the actual images demonstrate real-world effectiveness and support the validity of the simulation. In the first experiment, a software simulation based on the FLIR system model described in Section 5 demonstrates the effectiveness of the small kernel filter described in Section 4. In the second experiment, the small kernel filter is applied to actual FLIR images with good results. In the third experiment, the simulation software is used to examine tradeoffs in FLIR system design.

6.1. Simulation results

This section uses software simulation to compare the effectiveness of conventional reconstruction with the small-kernel restoration and reconstruction filters for use with FLIR systems. The imaging simulation software is based on the model described in Section 2, including the scene, the image acquisition PSF (with optics, detector, and scanning), sampling, noise, filtering, and post-filter reconstruction, and accounts for the FLIR components detailed in Section 5.

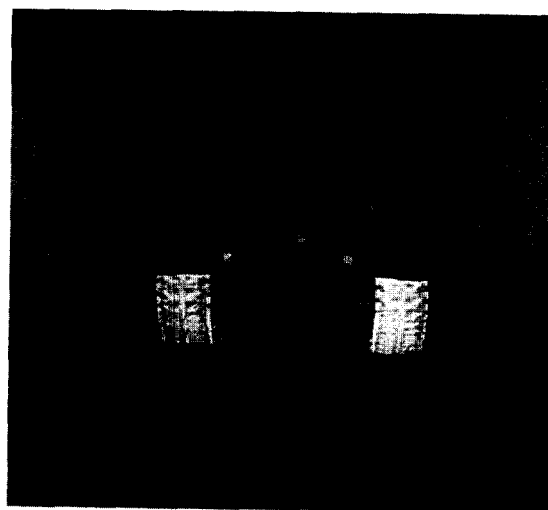


Fig. 2. Digital scene for simulation.

Table 1
Expected performance for the simulation

| Filter/reconstruction | Image fidelity | Structural content | Correlation quality | Coefficient correlation |
|-----------------------|----------------|--------------------|---------------------|-------------------------|
| Gaussian spot | 0.801 | 0.584 | 0.693 | 0.906 |
| Nearest neighbor | 0.810 | 0.660 | 0.735 | 0.905 |
| Bilinear | 0.817 | 0.605 | 0.712 | 0.914 |
| Cubic convolution | 0.840 | 0.651 | 0.746 | 0.924 |
| 3 × 3 small kernel | 0.884 | 0.884 | 0.884 | 0.940 |
| 5 × 5 small kernel | 0.890 | 0.890 | 0.890 | 0.943 |
| 7 × 7 small kernel | 0.891 | 0.891 | 0.891 | 0.944 |
| Unconstrained Wiener | 0.891 | 0.891 | 0.891 | 0.944 |

Software simulation allows assessment of overall system performance with precision, control, and flexibility that is not possible with real images [25].

The digital image in Fig. 2 is used as the ideal image of the scene. This digital scene was acquired with a thermoscope at about 315 m and digitized to 8 bits. The digital scene is 512×512 scene elements or *scenels*, with each scenel approximately 2.5 cm in the object plane (about 0.08 mrad). As described below, this digital scene is blurred with a simulated acquisition PSF, resampled, degraded by noise, and filtered and/or reconstructed. For these experiments, we used a value of $\bar{x}_0 = 0.5$ m in the object plane for the mean spatial detail. (This is 20 scenels or 1.6 mrad in the image in Fig. 2.) This value of \bar{x} is appropriate for human recognition processing of this

object [26]. The scene is not homogeneous (i.e., the level of spatial detail varies from one region to another). In particular, this value of \bar{x} is accurate for the region with the object, but is too small for the background regions in this image (which have greater spatial correlation). This is an appropriate value because, in practice, one would identify regions of interest (e.g., a region with potential targets) at low resolution and then process only the identified regions for high-resolution display or higher-level vision processing.

In this experiment, the simulated system has a sampling interval $\Delta = 36.6 \mu\text{m}$ (in the image plane) and focal length $D_i = 0.457$ m. (This is about 0.08 mrad/pixel.) With aperture $A = 0.203$ m and mean wavelength $\lambda = 10 \mu\text{m}$, the normalized OTF cutoff

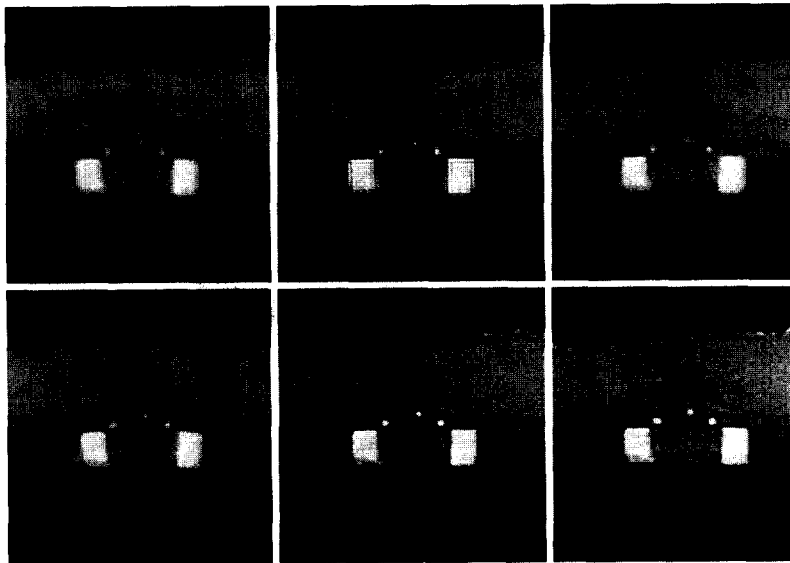


Fig. 3. Result images for simulation. Top row: Gaussian spot, nearest neighbor, bilinear. Bottom row: cubic convolution, 5 × 5 small kernel, unconstrained Wiener.

frequency is $\omega_0 = 1.63$ (about 20 cycles/mrad). We simulate a scanned system, with samples-per-dwell $P = 1$. For simulated imaging at $D_0 = 2500$ m, the scenel/pixel ratio is 8:1 in each dimension, sufficient so that the digital scene is effectively continuous with respect to the simulated imaging process. With an 8:1 scenel/pixel ratio, the 512×512 digital scene is sampled during simulated imaging to 64×64 and the mean spatial detail \bar{x} is 2.5 pixels (about 0.2 mrad). The simulation SNR is 32.

For the specific digital scene pictured in Fig. 2, the goal of the restoration and reconstruction processing is to take the 64×64 image simulating the system at 2500 m and produce the 512×512 image as it appears from 315 m. As described in Section 3, system performance can be measured by the fidelity of the output image r to the ideal image of the scene s . The expected performance measures for this system with four methods of conventional reconstruction, three variously sized small kernels, and unconstrained Wiener restoration are given in Table 1. The four conventional reconstruction methods (applied without filtering) are Gaussian spot, nearest neighbor, bilinear, and cubic convolution. The three small kernels have extent 3×3 pixels, 5×5 pixels, and 7×7 pixels and density $R = 8$ elements/pixel. (Note that the number of kernel elements is a function of both size and density, so for $R = 8$ the 3×3 kernel has 25×25 elements.) The last result is the unconstrained Wiener filter with density $R = 8$ and spatial extent equal to the image size 64×64 . As can be seen, the small kernels have significantly better expected performance than conventional reconstruction and perform nearly as well as the unconstrained optimal filter.

Resulting images for the different restoration and reconstruction methods are contained in Fig. 3. In

each case, the result image r has 8 samples/pixel to yield a 512×512 image. The nearest neighbor reconstruction gives the clearest illustration of the sampling density of the simulated FLIR system. Table 2 gives the computed quality measures for this particular image. The relative ordering for fidelity and correlation coefficient of the various restoration and reconstruction methods is as expected, but the performance measures are slightly greater than expected. This is because the observed autocorrelation for the whole image is slightly broader than the presumed autocorrelation. (Recall that the autocorrelation model parameter $\bar{x} = 2.5$ pixels, was set for the object which is more detailed than other regions of the scene.) The observed structural content and correlation quality is rather variable (and not as closely related to image quality as fidelity and the correlation coefficient). For these images, the structural content and correlation quality are somewhat larger than expected because the variance of the output image is larger than expected (due again to finer than expected detail).

In summary, these simulation results (in Tables 1 and 2) illustrate that images restored and reconstructed with the small kernels have greater fidelity than conventional reconstruction methods. Even the 3×3 kernel yields fidelity nearly as large as the optimal unconstrained Wiener filter. As can be seen in Fig. 3, the small kernel method yields an image with greater visual similarity to the scene.

6.2. Actual results

To test the actual performance of the restoration and reconstruction methods as a function of range, we acquired a sequence of FLIR images at various

Table 2
Computed performance for the simulation

| Filter/reconstruction | Image fidelity | Structural content | Correlation quality | Correlation coefficient |
|---------------------------|----------------|--------------------|---------------------|-------------------------|
| Gaussian spot | 0.946 | 0.857 | 0.902 | 0.974 |
| Nearest neighbor | 0.946 | 0.880 | 0.913 | 0.973 |
| Bilinear | 0.950 | 0.865 | 0.907 | 0.976 |
| Cubic convolution | 0.954 | 0.879 | 0.916 | 0.978 |
| 3×3 small kernel | 0.961 | 0.971 | 0.966 | 0.980 |
| 5×5 small kernel | 0.962 | 0.989 | 0.975 | 0.981 |
| 7×7 small kernel | 0.963 | 0.975 | 0.969 | 0.981 |
| Unconstrained Wiener | 0.963 | 0.969 | 0.966 | 0.981 |

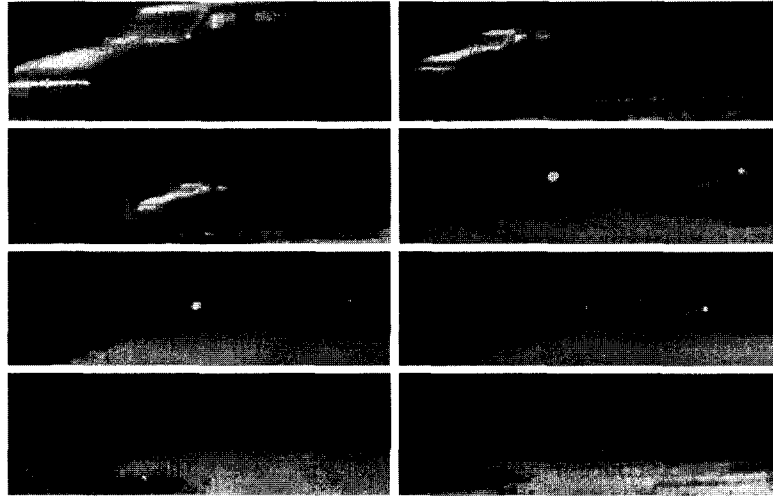


Fig. 4. Unprocessed actual images at distances of 100 m to 800 m at 100 m intervals.

distances while attempting to keep constant all variables other than range. The target object, an automobile, was imaged at distances from 100 m to 800 m at 100 m intervals. The images are pictured in Fig. 4. The confounding effects of the bright glare spot (caused by sunlight glint) in the 400 m, 500 m, and 600 m images on the results are discussed below. This experiment investigates the effectiveness of the methods at restoring the longer range images to the 100 m image (i.e., the longer-distance images are processed and then compared to the 100 m image).

The detector size of the staring FLIR system used to gather the image set is $\Delta = 31 \mu\text{m}$ (approximately 100% fill factor) and the focal length is $D_i = 100 \text{ mm}$, so at 100 m the pixel size is 3.1 cm. The mean spatial detail $\bar{x}_0 = 0.5 \text{ m}$ maps to 16 pixels at 100 m. The aperture is about $A = 43 \text{ mm}$ and the maximum wavelength is about $\lambda = 5.5 \mu\text{m}$, so the normalized OTF cutoff frequency is about $\omega_0 = 2.45 \text{ cycles/pixel}$. This is a two-dimensional array system, so there is no scanner, and with 100% fill, the detector size is the same size as the sampling interval. The theoretical diffraction-limited acquisition transfer function for this system is much broader than is actually observed in these images. In other words, the system caused significantly more blurring than the system specifications would indicate. This could be due to a combination of the optics, detector crosstalk, and other factors. Using the method in

[27], we estimated the actual PSF and set the defocus parameter ω_b in Eq. (33) to 0.281 in order to roughly fit the observed acquisition transfer function. Fig. 5 illustrates this defocus in the acquisition transfer function. From examination of the relatively uniform regions, we estimate the system SNR in these images at about 22.5.

Ideally, the only variable in the system for these images would be the range from the camera to the target automobile. The expected fidelity as a function of distance is given in Fig. 6. However, there are unavoidably some changes in the horizontal and vertical position of camera relative to the initial optical axis. Likewise, the imaging positions are not exactly at multiples of 100 m from the automobile. Fortunately, these XYZ differences are fairly easy to address. As the image is reconstructed, the transla-

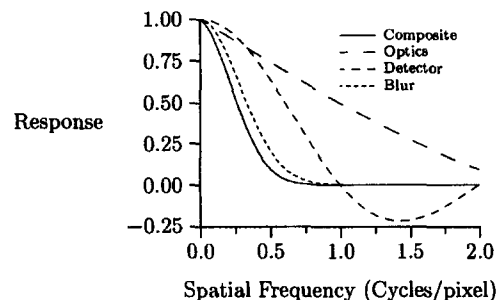


Fig. 5. Acquisition transfer function model for actual images.

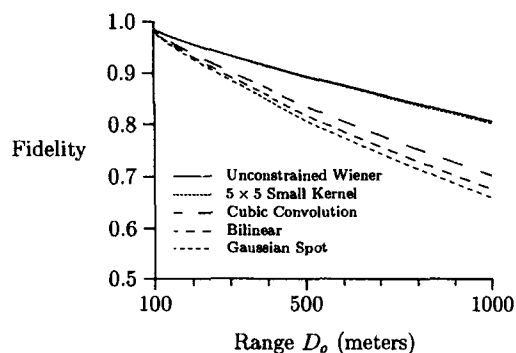


Fig. 6. Expected fidelity for actual images.

tion and scale are adjusted slightly to achieve the highest correlation between the result image and the 100 m image. (The small kernel restorations are adjusted slightly with post-filter bilinear interpolation.) Unfortunately, there are also slight differences in roll, pitch, and yaw from image to image. These are impossible to correct. There are other problems as well. For example, the 400 m, 500 m, and 600 m images each have a bright glare spot on the windshield that is not present in the 100 m image. Such variables are confounding, but despite these problems, the experiment provides meaningful results.

The correlation between the result images at the various distances and the 100 m image are reported in Table 3. In all cases where the scene did not have a bright glare spot (200 m, 300 m, 700 m, and 800 m), the images produced with the 5×5 small kernel are better correlated with the 100 m image than are the images reconstructed with conventional reconstruction. In the other images (400 m, 500 m, and 600 m), the glare is an additional degradation (relative to the 100 m image) that is not accounted for in the system model nor in the filter derivation. Because the small kernel is derived without accounting for glare noise, the small kernel filter sharpens the image and the glare too much. If the glare noise were included in the derivation, the small kernel filter would perform less sharpening (more blurring) and produce an image with greater fidelity.

Note that the image acquired at 300 m and restored with the 5×5 kernel is better correlated with the 100 m image than is the image acquired at 200 m and reconstructed with a Gaussian spot. Similarly, the image acquired at 700 m and restored with the 5×5 kernel is nearly as correlated with the 100 m image than is the image acquired at 300 m and reconstructed with a Gaussian spot. The result im-

Table 3

Computed performance for actual images

| Filter/reconstruction | Correlation coefficient | | | | | | |
|---------------------------|-------------------------|-------|-------|-------|-------|-------|-------|
| | 200 m | 300 m | 400 m | 500 m | 600 m | 700 m | 800 m |
| Gaussian spot | 0.915 | 0.847 | 0.548 | 0.544 | 0.691 | 0.709 | 0.681 |
| Bilinear | 0.933 | 0.896 | 0.555 | 0.552 | 0.726 | 0.800 | 0.763 |
| Cubic convolution | 0.937 | 0.888 | 0.554 | 0.547 | 0.729 | 0.807 | 0.757 |
| 5×5 small kernel | 0.946 | 0.926 | 0.536 | 0.545 | 0.717 | 0.841 | 0.802 |

Note: the 400 m, 500 m, and 600 m images are degraded by glare.

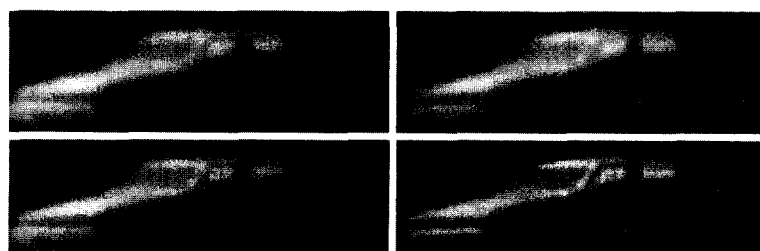


Fig. 7. Processed actual images at 200 m. Top row: Gaussian spot and bilinear. Bottom row: cubic convolution and 5×5 small kernel.

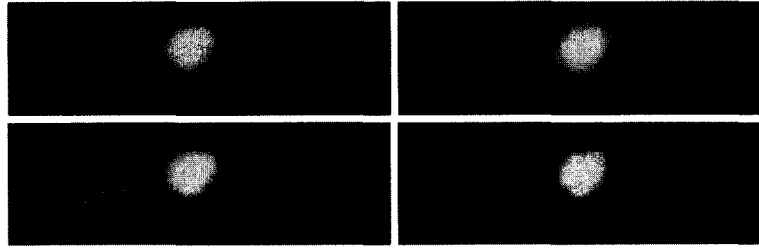


Fig. 8. Processed actual images at 400 m. Top row: Gaussian spot and bilinear. Bottom row: cubic convolution and 5×5 small kernel.

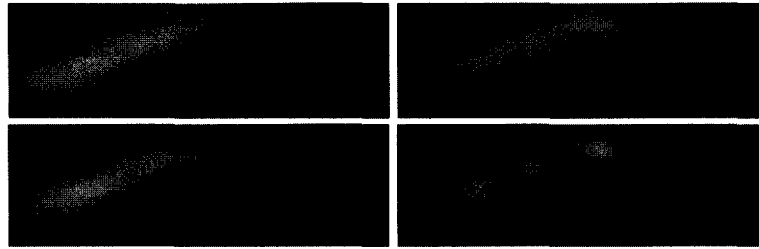


Fig. 9. Processed actual images at 800 m. Top row: Gaussian spot and bilinear. Bottom row: cubic convolution and 5×5 small kernel.

ages for 200 m, 400 m, and 800 m are pictured in Figs. 7–9. These images also illustrate the improvements in the images restored with the small kernel.

6.3. System design

The final experiment presented in this paper, explores the effects of design tradeoffs on system performance. Specifically, the mean spatial detail (as a function of range), the presampling acquisition blur (as a function of the cutoff of the diffraction limited optical system), and the SNR are varied and the

expected fidelity is reported for each configuration. The results are obtained using the software simulation described above.

The scene mean-spatial-detail in the object plane is fixed at 0.5 m and three ranges are considered: 315 m, where the image-plane mean-spatial-detail is about 1.6 mrad; 1250 m, where the image-plane mean-spatial-detail is about 0.4 mrad; and 5000 m, where the image-plane mean-spatial-detail is about 0.1 mrad. To simplify the consideration of blurring, the acquisition system is modeled with only diffraction-limited optics, detector with 100% fill factor,

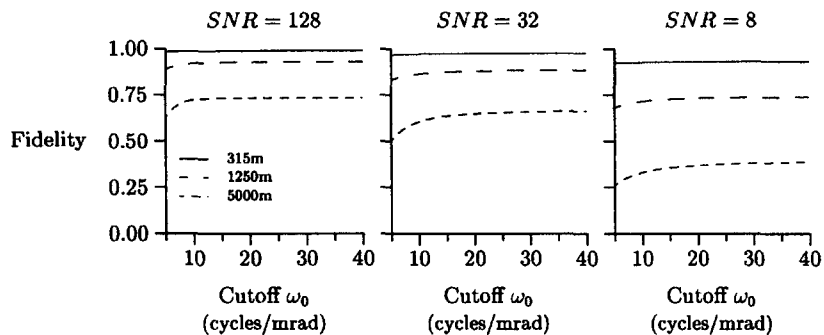


Fig. 10. Fidelity for system evaluation.

and noise (i.e., no defocus blur or scanner), with the OTF cutoff expressed in cycles/mrad. The sampling interval is fixed at 0.1 mrad, a figure appropriate for narrow-field-of-view tactical FLIRs. Three SNRs, 128, 32, and 8, are considered.

The results of the simulations are presented in Fig. 10. For the high SNR images, restoration processing is able to correct for a fairly high degree of presample blurring, even in images at great distance. At 5000 m, where the mean spatial detail is equal to the sample spacing of 0.1 mrad, in the images with SNR = 128, performance is good even with the acquisition transfer function cutoff down to about 10 cycles/mrad, which is equal to the sampling frequency. Restoration is less successful in correcting presample blurring in images with low SNR. This sort of simulation could be used in the design of a FLIR system, for example to choose the optical configuration based on anticipated distance and SNR. Of course, in practice other more extensive simulations would be used (e.g., with SNR a function of distance to account for atmospheric attenuation [28]).

A benefit of restoration processing is that it allows greater freedom in design tradeoff between blurring and aliasing. Without restoration, the utility of presample blurring to reduce aliasing is usually limited in the system design by the level of blurring that is visually acceptable. Of course, this may not maximize the total information content of the image that is passed to the human or ATR [29]. Restoration processing can compensate for blurring and so allows greater latitude in designing the system to reduce aliasing.

7. Conclusion

This paper describes the design of an efficient filter that significantly improves the performance of FLIR imaging systems. The filter is based on a comprehensive model of the digital imaging process that accounts for the significant effects of sampling and reconstruction as well as acquisition blur and noise. The filter both restores, partially correcting degradations introduced during image acquisition, and interpolates, increasing apparent sharpness and improving reconstruction. The filter derivation is conditioned on explicit constraints on spatial support

and density so that it can be implemented efficiently and is practical for real-time applications. Subject to these implementation constraints, the filter optimizes end-to-end system fidelity, yielding an image that more closely resembles an 'ideal' close-range image. In experiments with simulated and actual FLIR systems, the filter significantly increases fidelity and apparent clarity with relatively little computation.

References

- [1] J.D. Howe, in: *Infrared and Electro-Optical Handbook*, ed. M. Dudzik, Vol. 4 (ERIM and SPIE, 1993) ch. 2, pp. 55–120.
- [2] W.F. Schreiber, *Fundamentals of Electronic Imaging Systems: Some Aspects of Image Processing* (Springer-Verlag, New York, 1986).
- [3] H.C. Andrews and B.R. Hunt, *Digital Image Restoration* (Prentice-Hall, Englewood Cliffs, NJ, 1977).
- [4] F.O. Huck, C.L. Fales, N. Halyo, R.W. Samms and K. Stacy, *J. Opt. Soc. Am. A* 2(10) (1985) 1644–1666.
- [5] C.L. Fales, F.O. Huck, J.A. McCormick and S.K. Park, *J. Opt. Soc. Am. A* 5(3) (1988) 300–314.
- [6] J.A. Ratches, *Opt. Eng.* 15(6) (1976) 525–530.
- [7] W. McCracken and L. Wajsfelner, in: *Thermal Imaging*, Proc. SPIE 636 (1986) pp. 31–35.
- [8] H.V. Kennedy, *Opt. Eng.* 30(11) (1991) 1771–1778.
- [9] S.E. Reichenbach and S.K. Park, *IEEE Trans. Signal Process.* 39(10) (1991) 2263–2274.
- [10] S.K. Park and R.A. Schowengerdt, *Comput. Vision Graphics Image Process.* 23 (1983) 258–272.
- [11] F. Rosell and G. Harvey (eds.), *The Fundamentals of Thermal Imaging Systems*, NRL Report 8311, ADA 073 763 (1979).
- [12] J.A. Ratches, W.R. Lawson, L.P. Obert, R.J. Bergemann, T.W. Cassidy and J.M. Swenson, *Night Vision Laboratory Static Performance Model for Thermal Viewing Systems*, ECOM-7043 (1975).
- [13] D.L. Schumaker, J.T. Wood and C.D. Thacker, *FLIR Performance Handbook*, DCS Corp. (1988).
- [14] *Night Vision and Electronic Sensors Directorate Report, FLIR92 Thermal Imaging Systems Performance Model*, US Army, January (1993).
- [15] E.H. Linfoot, *J. Opt. Soc. Am.* 46(9) (1956) 740–752.
- [16] W.K. Pratt, *Digital Image Processing*, 2nd Ed. (John Wiley and Sons, New York, 1991).
- [17] Y. Itakura, T. Suteo and T. Takagi, in: *Infrared Physics*, Vol. 14 (Pergamon Press, Elmsford, NY, 1974) pp. 17–29.
- [18] K. Ben-Yosef, N. Wilner, S. Simhony and G. Feigin, *Appl. Opt.* 24(14) (1985) 2109–2113.
- [19] H.H. Hopkins, *Proc. R. Soc. London A* 231 (1955) 91–103.
- [20] P.A. Stokseth, *J. Opt. Soc. Am.* 59(10) (1969) 1314–1321.
- [21] J.E. Murguia, J.M. Mooney and W.S. Ewing, *Opt. Eng.* 29(7) (1990) 786–794.

- [22] R.D. Hudson, Jr., *Infrared System Engineering* (John Wiley and Sons, New York, NY, 1969).
- [23] W.L. Wolfe and G.J. Zissis (eds.), *The Infrared Handbook*, ERIM, Ann Arbor, MI (1985).
- [24] J.D. Vincent, *Fundamentals of Infrared Detector Operation and Testing* (John Wiley and Sons, New York, NY, 1990).
- [25] S.E. Reichenbach, S.K. Park, R. Alter-Gartenberg and Z. Rahman, in: *Stochastic and Neural Methods in Signal Processing, Image Processing and Computer Vision*, Proc. SPIE 1569 (1991) pp. 422–433.
- [26] J. Johnson, in: *Image Intensifier Symposium*, U.S. Army Research and Development Laboratories (1958) pp. 249–273.
- [27] S.E. Reichenbach, S.K. Park and Ramkumar Narayanswamy, *Opt. Eng.* 30(2) (1991) 170–177.
- [28] S.E. Reichenbach, S.K. Park, G.F. O'Brien and J.D. Howe, in: *Visual Information Processing*, Proc. SPIE 1705 (1992) pp. 165–176.
- [29] J.A. McCormick, R. Alter-Gartenberg and F.O. Huck, *J. Opt. Soc. Am. A* 6(7) (1989) 987–1005.