Image-based prediction of imaging and vision performance

Saar Bobrov* and Yoav Y. Schechner

Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel
*Corresponding author: saarbob@rafael.co.il

Some scenarios require performance estimation of an imaging or a computer vision system prior to its actual operation such as in system design, as well as in tasks of high risk or cost. To predict the performance, we propose an image-based approach that accounts for underlying image-formation processes while using real image data. We give a detailed description of image formation from scene photons to image gray levels. This analysis includes all the optical, electrical, and digital sources of signal distortion and noise. On the basis of this analysis and our access to the camera parameters, we devise a simple image-based algorithm. It transforms a baseline high-quality image to render an estimated outcome of the system we wish to operate or design. We demonstrate our approach on thermal imaging systems (infrared spectrum, $3-5$ μm). © 2007 Optical Society of America

OCIS codes: 100.2000, 110.3080, 110.2960, 110.3000.

1. INTRODUCTION

Some scenarios require performance estimation of an imaging or a computer vision system prior to its operation. One scenario is system design, which requires setting of various optical, electronic, and algorithmic specifications. There is a need to know how settings affect the system output and consequently how the system design should be altered. Another scenario is high-risk operation, where actual operation should preferably be avoided unless a successful outcome is likely. Such is the case with disposable systems such as space probes or guided missiles as well as with intrusive medical imaging operations. Thus, our work focuses on estimating the output of imaging and computer vision systems as if they view typical scenes, without actual operation. The contributions of this paper are as follows:

1. Introduction of the problem, as described above: pointing out cases and a need for forecasting an image and performance of subsequent computer vision prior to actual acquisition by an intended system.
2. Introduction of an effective solution to the problem.

System analysis has used physics-based methods. It attempts prediction simply by looking at numeric specifications of the imager [1–3] and couples them to various figures of merit [4–7]. Another physics-based method simulates synthetic scenes and then renders them based on the sensor specifications [8–10]. These methods may face difficulty in accounting for the complexity and unpredictability of real-world scenes. This difficulty exists particularly beyond the visible spectrum, where computer graphics models have not matured.

A related matter is image-based rendering. It may render the output of the imager in question. Here the input is high quality (HQ) images of real scenes taken by a dedicated system. These images can be aberrated to render the output, which is particularly useful when the system we need to design or operate is of low quality (LQ). Indeed, sensors used in disposable systems or intrusive instruments often have LQ due to space, weight, and price constraints. Related image-based methods (termed example-based in this context) include texture synthesis [11–13] and image analogies [14–16]. These methods currently require learning of the cross-modal transformation every time; they are iterative and thus computationally complex. Moreover, they might yield unsatisfying results, since the operations they perform do not always match the operations performed by the systems.

To counter the drawbacks of previous methods, we propose a hybrid approach. It is image based, but it exploits knowledge of the physical and electronic processes occurring in imaging systems. Hence, we use a HQ system to measure the scene and then use a physics-based algorithm for transforming the HQ image into an estimated output, of the LQ system. We analyze the fundamental processes and the inner workings of both systems, producing a unified model of an imaging process that is applicable to a wide variety of imaging modalities. This model is used for image-based rendering and subsequently for grading a computer-vision task.

As a case study, we examine a thermal imaging seeker mounted in a missile. Hence, some of the optical effects we address apply in general to thermal imaging [1,17–19]. Being on a disposable missile and constrained by price, space, and weight, this imaging system has LQ. This system cannot be operated at will: It is cooled by unrecycled gas, and hence once the gas is depleted (a short time after activation) the system becomes useless. As a result, it is preferable to avoid activation before there is a prediction that this LQ system can lock on targets in the

scene. We perform this estimation using an accompanying thermal sight. The thermal sight is an HQ system, since it is not disposable and not compromised by the mentioned constraints. We use these systems to demonstrate this approach in experiments. In addition to visual demonstration of the rendering, we study how the transformed image helps in predicting the system's performance, such as in a lock-on task.

2. SCENE PHOTONS TO A GRAY-LEVEL IMAGE

This section explains the evolution of the scene photons, from the stage at which they leave the scene and enter into the imaging system until their representation as a gray-level image. Some parts of this path have often been described [20–23] in isolation from the other parts. Here we present the complete path.

An imaging system is commonly composed of an optical assembly, a detector module, and signal processing electronics. The imaging process can be described as a functional flow diagram, as depicted in Fig. 1. The input to any imaging system is a two-dimensional (2D) scene radiance $L_\lambda^{\mathrm{scn}}(x,y)$ in units of [photon/cm^2 μm sr s]. Here (x,y) are the spatial coordinates and λ is the wavelength. The optical assembly gathers the scene radiance onto the detector array. The detector module converts the radiation (photons) into an electrical signal, which enters a signal processing unit. This unit enhances the signal and converts it into the output image of the particular system. A typical output is an 8 bit image matrix $I_{\mathrm{out}}(m,n)$ [gray level]. Note that we use (x,y) to express continuous spatial coordinates and (m,n) to express discrete row and column coordinates.

In this section we describe each process in context. As an example, we sometimes refer to two types of thermal imaging systems. The HQ system is a staring system, which uses a 2D detector array. The LQ system is a scanning system having a one dimensional (1D) vertical detector column array. This vertical column array sweeps the scene horizontally.

A. Optical Assembly
An optical assembly induces all the effects that occur between the outer world (scene radiance) and the detector. This assembly is commonly built of focusing optics, possibly a scanning mirror, a filter, and a cold shield. The focusing optics (typically lenses) collect the scene radiance and project it onto the detector plane. In scanning systems, a mirror sequentially scans the detector swath over the scene to produce a 2D image. The scanning mirror is used in conjunction with column detector arrays. The filter specifies the spectral range for imaging. In thermal imaging this filter is cooled so that it practically does not
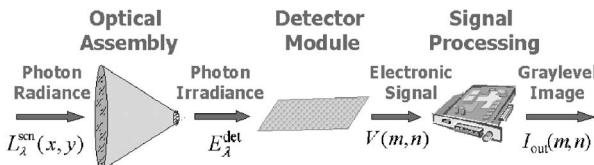
Fig. 1.　Functional flow diagram of an imaging system.
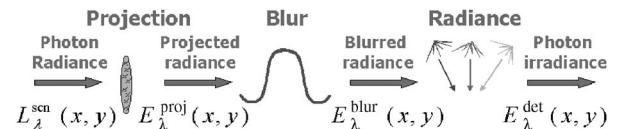
Fig. 2.　Functional flow diagram of an optical assembly.
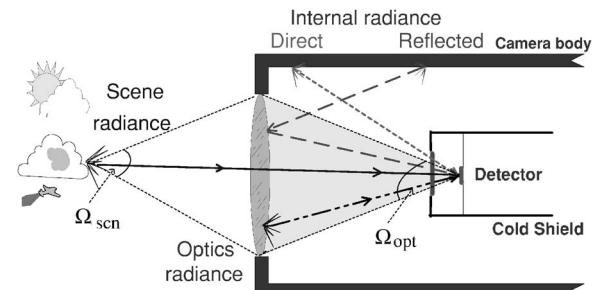
Fig. 3.　Different contributions to the photon irradiance on the detector. Tracks of different contributions are in different line styles.

radiate (hence, it is termed a cold filter). In addition, in thermal imaging, the detector might sense radiation emitted from the body of the system. This radiation is obstructed from reaching the detector, using a cold shield. As depicted in Fig. 2, the effects of the optical assembly can be decomposed as projection, blur, and ambient radiance processes. In the following, we explain each subprocess.

1. Radiance Projection and Blur
The main role of an optical assembly is to project the scene radiance onto the detector. As depicted in Fig. 3, the scene radiance is gathered through a solid angle $\Omega_{\mathrm{scn}} \simeq A_{\mathrm{opt}}/R^2$ [sr], where A_{opt} is the area of the optical aperture and R is the distance to the scene. The radiance $L_\lambda^{\mathrm{scn}}(x,y)$ is then projected onto the detector with a magnification ratio $a=f_{\mathrm{opt}}/R$, where f_{opt} is the focal length of the optics. The solid angle set by the optics toward a detector element is $\Omega_{\mathrm{opt}} \simeq A_{\mathrm{opt}}/f_{\mathrm{opt}}^2 = (1/a^2)\Omega_{\mathrm{scn}}$ [sr]. The scene contribution to the irradiance on the detector is thus

$$E_\lambda^{\mathrm{proj}}(x,y) = \Omega_{\mathrm{opt}} L_\lambda^{\mathrm{scn}}\left(\frac{x}{a},\frac{y}{a}\right)\tau_{\mathrm{opt}}(\lambda)\left[\frac{\mathrm{photon}}{\mathrm{cm}^2 \cdot \mu\mathrm{m} \cdot \mathrm{s}}\right], \quad (1)$$

where $\tau_{\mathrm{opt}}(\lambda)$ is the transmittance of the optics.

The projection is not perfect, since the optical elements create blur. This is expressed as a convolution with a point spread function (PSF), $h_{\mathrm{opt}}(x,y)$

$$E_\lambda^{\mathrm{blur}} = h_{\mathrm{opt}} * E_\lambda^{\mathrm{proj}}, \quad (2)$$

where $E_\lambda^{\mathrm{proj}}$ is given by Eq. (1). For scanning systems, $h_{\mathrm{opt}}(x,y)$ includes scan blur, $h_{\mathrm{scan}}(x)$ caused by horizontal motion with velocity v_{scan} during the integration time t_{int},

$$h_{\mathrm{scan}}(x) = \mathrm{rect}\left(\frac{x}{v_{\mathrm{scan}}t_{\mathrm{int}}}\right). \quad (3)$$

2. Ambient Radiance
Systems may sense a radiance component that is not directly related to the scene in view. In thermal imaging, this radiance comes from the camera body and from the optics. Figure 3 depicts the different radiance contributions to the photon irradiance on the detector. The camera body is hot, and thus it radiates. While most of this radiation is blocked by a cold shield, part of it may reach the detector directly, contributing E_λ^{body} to the photon irradiance. In addition, some of the camera body radiance is reflected by the optics toward the detector, contributing E_λ^{rfl} to the photon irradiance. Furthermore, the optics is hot and radiating, thus contributing E_λ^{opt} to the photon irradiance. All the above-mentioned contributions add up to the total spectral photon irradiance [24,25] beyond E_λ^{blur} derived in Eq. (2),

$$E_\lambda^{\text{det}}(x,y) = E_\lambda^{\text{blur}}(x,y) + E_\lambda^{\text{body}} + E_\lambda^{\text{rfl}} + E_\lambda^{\text{opt}}, \qquad (4)$$

in units of $[\text{photon}/\text{cm}^2/\mu\text{m}/\text{s}]$. To calculate these radiance contributions we use the temperature of the camera body T_b and the Planck equation [22,24,25]:

$$L_\lambda(T_b) = \frac{c_3}{\pi\lambda^4(e^{c_2/\lambda T_b} - 1)}\left[\frac{\text{photon}}{\text{cm}^2 \cdot \mu\text{m} \cdot \text{sr} \cdot \text{s}}\right], \qquad (5)$$

where $c_2 = 14388\,\mu\text{m}\,°\text{K}$, $c_3 = 1.8837 \times 10^{23}\,\mu\text{m}^3/(\text{cm}^2 \cdot \text{s})$. Normally T_b is near the ambient temperature (which is much hotter than the detector arid cold shield), while in high-speed missile flight it may be much hotter. More details are given in Appendix A.

B. Detector Module
In Subsection 2.A we have discussed the photon irradiance on the detector plane. A detector converts the photon irradiance [Eq. (4)] into an electrical signal. As depicted in Fig. 4, the detector function can be decomposed into photodetection, cross talk, sampling, noise, and readout mechanisms. In the following, we describe each process.

Our systems use photovoltaic detectors. Electrons are excited in response to photon irradiance. They are accumulated on a capacitor during an integration time. After integration, the voltage over the capacitor is read. Photodetection expresses the excitation rate of electrons in the detector material in response to the photon irradiance for quantum efficiency $\eta_Q(\lambda)[e/\text{photon}]$; this rate is

$$\dot{N}_e^{\text{gen}}(x,y) = \int_\lambda E_\lambda^{\text{det}}(x,y)\,\eta_Q(\lambda)\mathrm{d}\lambda + \frac{i_{\text{dark}}}{qA_d} \qquad (6)$$

in units of $[e/\text{cm}^2 \cdot \text{s}]$, where $E_\lambda^{\text{det}}(x,y)$ is given in Eq. (4). Here, i_{dark} is the dark current at a detector element, q is the electron charge, and A_d is the effective area of each detector element.

Photogenerated electrons may drift across the detector array, creating cross talk. This is expressed as a convolution of $\dot{N}_e^{\text{gen}}$, defined in Eq. (6), with a 2D PSF (a 1D PSF for a 1D column detector array) $h_{\text{xtk}}(x,y)$

$$\dot{N}_e(x,y) = h_{\text{xtk}} * \dot{N}_e^{\text{gen}}\left[\frac{e}{\text{cm}^2 \cdot \text{s}}\right]. \qquad (7)$$

This PSF can be provided by the manufacturer or it can be precalibrated in the laboratory in the detector level.

At this stage, we have an instantaneous distribution of charges on the surface of the detector array. This distribution is integrated in each detector element area A_d and in time t_{int} (integration time). The integrated signal is sampled by the array pitch $(\Delta x_d, \Delta y_d)$,

$$N_e^{\text{samp}}(m,n) = t_{\text{int}} \int \int_{A_d} \dot{N}_e(x,y)\mathrm{d}x\mathrm{d}y\,[e], \qquad (8)$$

where $\dot{N}_e$ is given in Eq. (7). This operation can be expressed as

$$N_e^{\text{samp}}(m,n) = A_d t_{\text{int}}[h_{\text{det}} * \dot{N}_e]_{(x_n y_m)}\,[e], \qquad (9)$$

where the PSF of the detector element shape $h_{\text{det}}(x,y)$ is a rectangular window of size $\alpha \times \beta\,[\mu\text{m}]$

$$h_{\text{det}}(x,y) = \text{rect}(x/\alpha, y/\beta). \qquad (10)$$

Equation (9) also expresses sampling by the array pitch $(\Delta x_d, \Delta y_d)$:

$$x_n = (n - 1/2)\Delta x_d, \quad y_m = (m - 1/2)\Delta y_d. \qquad (11)$$

We now have samples of charge $N_e^{\text{samp}}(m,n)$ collected in each detector element. This charge is linearly translated into a voltage $V(m,n)$ by

$$V_{\text{read}}(m,n) = qN_e^{\text{samp}}(m,n)/C\,[\text{V}], \qquad (12)$$

where C is the capacitance of the readout capacitor. In some systems, there is a readout capacitor for each detector element. In others it is common to a set of elements (in
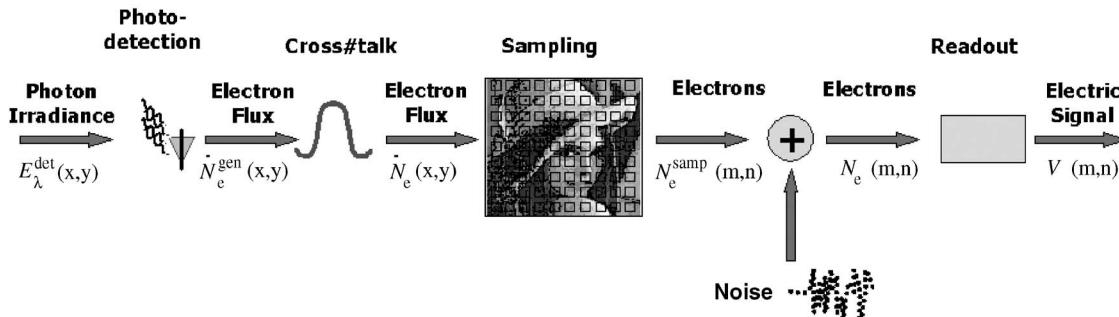
Fig. 4.   Flow diagram of a detector module.
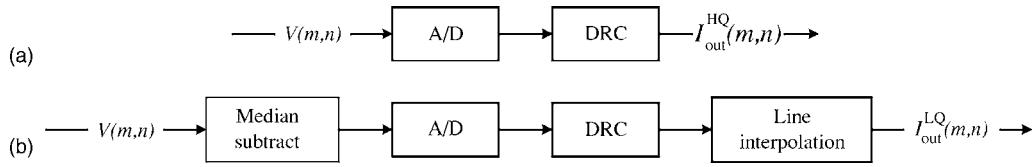
Fig. 5.   Image-processing flow diagram (a) in the HQ system, (b) in the LQ system.

a CCD there is typically a single readout capacitor, which processes the whole array).

Detection is a noisy process. This is modeled as an addition to Eq. (12) of a white-noise matrix $V_{noise}(m,n)$ with standard deviation (STD) ν_{noise}:

$$V(m,n) = V_{read}(m,n) + V_{noise}(m,n) \; [V]. \qquad (13)$$

This noise is composed of shot noise, spatial noise, and excess noise. The shot noise ν_{shot} has STD [20,22,26]

$$\nu_{shot} = \frac{q}{C} \sqrt{N_e^{samp}} \; [V]. \qquad (14)$$

In addition, spatial noise expresses the nonuniformity of the detector array. This nonuniformity is precalibrated and is compensated for in a postprocess, called nonuniformity correction (NUC). Nevertheless, NUC is not perfect, and some residual nonuniformity (RNU) remains. This is expressed as spatial noise with STD ν_{RNU} [$volt$]. There are additional noise sources, including readout amplifier noise and digital quantization noise. Their overall STD is expressed as ν_{excess} [$volt$]. The total noise STD is thus

$$\nu_{noise} = \sqrt{\nu_{shot}^2 + \nu_{RNU}^2 + \nu_{excess}^2} \; [V]. \qquad (15)$$

C. Internal Signal Processing Module
An internal signal processing module converts the readout voltage on the capacitor $V(m,n)$ into the output image of the imaging system $I_{out}(m,n)$. In contrast to previous processes, this process is unique to each system brand. Figure 5 shows the flow diagram of this module for the LQ arid the HQ systems that we examine.

Most systems use standard, linear analog-to-digital conversion (ADC) and dynamic range compression (DRC) for display. The ADC for b bits in our systems is linear for $V \in [V_{min}, V_{max}]$, where V_{min} and V_{max} are constants of each system brand:

$$I_b(m,n) = \left[(2^b - 1)\frac{V(m,n) - V_{min}}{V_{max} - V_{min}}\right]. \qquad (16)$$

Values beyond or above that are assigned 0 or $(2^b - 1)$ respectively. The values in $I_b(m,n)$ are quantized.

As for the DRC, our systems use a linear global gain, g, and a global offset, o, to achieve the conversion to 8 bits for display:

$$I_{out}(m,n) = [gI_b(m,n) + o]. \qquad (17)$$

The parameters g and o are determined in real time so that the histogram is stretched over the dynamic range of the display. In addition to the ADC and the DRC, the particular LQ system we use employs median subtraction prior to the ADC, as explained below. It also employs line interpolation after the DRC to achieve a standard output

format $I_{out}(m,n)$. The detector has only 120 elements; therefore it outputs 120 rows. To achieve standard National Television Systems Committee field format (1 field =240 rows) a new row is created between each two rows, that is the average of the two rows.

Median subtraction in scanning systems. The thermal readout $V(m,n)$ given in Eq. (13) has a relatively large dc contribution when imaging an outdoor scene. This is because in most natural terrestrial scenes, the intrascene temperature variations are small: $\mathcal{O}(1°K)$ relative to a dc temperature of $\mathcal{O}(300°K)$. It is beneficial to omit the dc of the continuous signal to make the subsequent ADC more effective, as ADC would concentrate only on the variations. For this reason, scanning thermal imagers often use [27,28] an analog median subtraction as described below.

The LQ system we work with has a vertical array of detector elements, which scans the 2D scene horizontally. Each detector element produces a scanned signal as it traces a row. The median value of each row is subtracted from the readout values of that row by an analog circuit. We note that the data are digitized temporarily to calculate the median value. Then the subtraction is implemented by an analog circuit over the continuous signal. Consider row m with n_{max} elements. The voltage in that row is

$$\mathbf{V}_m = [V(m,1), V(m,2), \ldots V(m,n_{max})], \qquad (18)$$

where $V(m,n)$ is derived in Subsection 2.B. The median of this set of pixels is

$$d_m = \mathrm{median}(\mathbf{V}_m). \qquad (19)$$

The median subtraction operation is then

$$\mathbf{V}^{med} = \begin{bmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_M \end{bmatrix} - \begin{bmatrix} d_1 \\ \vdots \\ d_M \end{bmatrix}, \qquad (20)$$

where M is the number of matrix rows. Hence, for our LQ system we use $\mathbf{V}^{med}$ instead of $V(m,n)$ in Eq. (16).

3. CROSS-SENSOR TRANSFORMATION

As explained in Section 1, we wish to predict the performance of an LQ system, using an HQ image. We do it by simulating an LQ image I_{out}^{LQ} by transforming an input HQ image I_{out}^{HQ}. This transformation is based on the parameters of the HQ system. These include temporally changing parameters (which are extracted from the system with the video output) such as DRC parameters (g and o), integration time t_{int}^{HQ}, and body temperature T_b^{HQ}. Other included parameters are constants of each system (which are specified by the manufacturer or precalibrated in laboratory) such as focal length, detector pitch, amplifier

capacitance, and optical parameters. Here, we assign an LQ or HQ label to some variables to distinguish between the systems.

The transformation starts by inverting the operations of the HQ system. This yields an estimate of the scene radiance $\hat{L}^{\text{scn}}$ [photon/pixel/s]. We may then apply the operations of the LQ system. This yields an estimation of the LQ image, $\hat{I}_{\text{out}}^{\text{LQ}}(m,n)$.

A. Overcoming Some Pitfalls
We need to overcome some pitfalls in the transformation process. Some processes described in Section 2 operate on continuous signals, whereas the HQ data is already sampled and quantized. Special attention should be paid to inversion of blur operations as in Eqs. (2), (7), and (9), since deblurring is an unstable operation. In addition, any HQ image includes noise and aliasing, which cannot be inverted. In the following, we address these pitfalls.

1. Overcoming Deblurring Instability
A transformation that inverts the HQ operations estimates the scene radiance. However, we are not interested in the estimation of the scene radiance: This is just an intermediate stage to estimate the LQ image. By definition, an LQ system has lower optical and detector qualities than those of an HQ system. For this reason, as depicted schematically in Fig. 6, the blur operators of the LQ system (H^{LQ} in the frequency domain) are more band limited than those of the HQ system (H^{HQ}); i.e., the cutoff frequency of the LQ system is lower than that of the HQ system. Therefore, unstable frequency components in a deblurred $I_{\text{out}}^{\text{HQ}}$ are nulled in the overall process. Thanks to this observation, we null the values corresponding to frequencies outside the passband of the LQ system before applying deblur operations to the HQ data. To demonstrate the difference between the blur operators of the systems, Fig. 7 (below) depicts the optical PSFs of the HQ
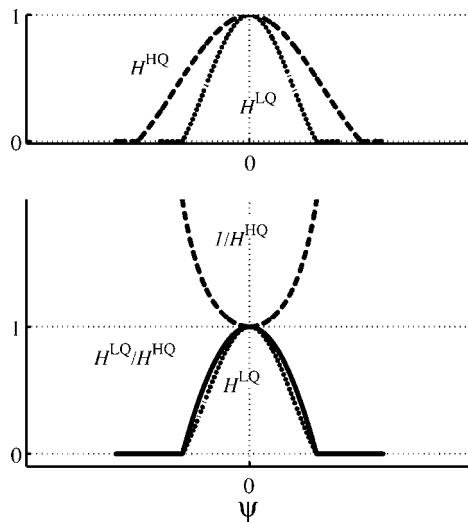
Fig. 6. Stability of a deblurring operation. Top, schematic plots of the blur frequency responses of an LQ system and an HQ system. Bottom, a schematic plot of an inversion of an HQ blur operation, proceeded by an LQ blur operation. On its own, inversion of H^{HQ} is unstable at high frequencies. However, subsequent application of H^{LQ} results in a stable operation.
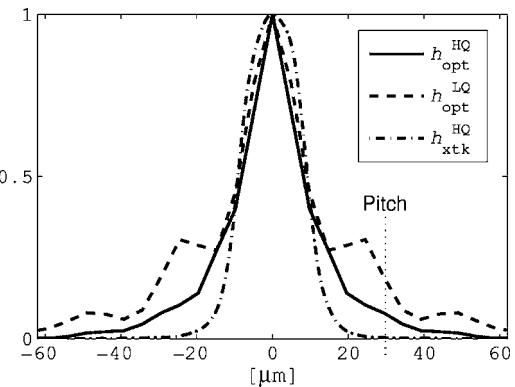
Fig. 7. Optical and cross-talk PSFs. The figure presents 1D profiles of $h_{\text{opt}}^{\text{HQ}}$, $h_{\text{opt}}^{\text{LQ}}$ and $h_{\text{xtk}}^{\text{HQ}}$, as well as the horizontal pitch of both systems (30 μm). The optical blur of the LQ system is stronger than that of the HQ system.

system and the LQ system.

2. Sampled Data
Our data have already been sampled by the HQ detector. We do not reconstruct the continuous signal. This means that analog blur operations such as optical blur, cross talk, and detector spatial integration should be matched to the sampling dimensions ($\Delta x_{\text{d}}, \Delta y_{\text{d}}$). Hence, we use

$$\hat{h}(m,n) = h(m\Delta y_{\text{d}}, n\Delta x_{\text{d}}), \qquad (21)$$

where h is a general continuous blur PSF as in Eqs. (2), (7), and (9).

3. Noise
Noise is an additional issue. When we invert the HQ operations, we cannot invert the noise addition, due its randomness. However, we found a way to circumvent this problem. Let us consider an arbitrary signal s with noise STD ν. Suppose we wish to render the signal $\hat{s}$, whose noise STD is $\hat{\nu} > \nu$. There is no need to denoise the signal s. Rather, we should add some noise to simulate $\hat{s}$. Explicitly, we should add noise with STD of $\sqrt{\hat{\nu}^2 - \nu^2}$. Our case is similar: Our goal is to render an LQ image with realistic noise rather than denoising an HQ image. The noise STD of an LQ image is larger than that of an HQ image. Therefore, it is possible to add noise to the estimated LQ image such that its STD is equal to a true LQ image noise.

To achieve this, we first calculate the HQ noise STD, $\nu_{\text{noise}}^{\text{HQ}}$, using Eq. (15). For a moment, we ignore this noise during the cross-sensor transformation (i.e., we do not remove the noise). As we go along the transformation, we concatenate all the factors that multiply the image (detector element area, integration time, etc.),

$$\kappa = \frac{C^{\text{HQ}} A_{\text{d}}^{\text{LQ}} t_{\text{int}}^{\text{LQ}} \bar{\eta}_{\text{Q}}^{\text{LQ}} \Omega_{\text{opt}}^{\text{LQ}} \tau_{\text{opt}}^{\text{LQ}}}{C^{\text{LQ}} A_{\text{d}}^{\text{HQ}} t_{\text{int}}^{\text{HQ}} \bar{\eta}_{\text{Q}}^{\text{HQ}} \Omega_{\text{opt}}^{\text{HQ}} \tau_{\text{opt}}^{\text{HQ}}}. \qquad (22)$$

The factor κ expresses the amplification that $\nu_{\text{noise}}^{\text{HQ}}$ undergoes in the cross-sensor transformation. When noise is added to the estimated LQ image, we do not use $\nu_{\text{noise}}^{\text{HQ}}$ as the noise STD. Rather, we use

$$\hat{\nu}_{\text{noise}}^{\text{LQ}} = \sqrt{(\nu_{\text{noise}}^{\text{LQ}})^2 - (\kappa \nu_{\text{noise}}^{\text{HQ}})^2} \qquad (23)$$

to subtract the HQ noise effects. As explained above, this operation is possible, thanks to the fact that

$$\nu_{\text{noise}}^{\text{LQ}} > \kappa \nu_{\text{noise}}^{\text{HQ}}. \qquad (24)$$

4. Aliasing
Aliasing may be created in the sampled image $V(m,n)$ if the continuous image $\dot{N}_{\text{e}}(x,y)$ contains spatial frequencies that are higher than the sampling frequency of the detector. Aliasing in the HQ image cannot be inverted, if it occurs. When the sampling frequency of the system surpasses the Nyquist frequency dictated by the optics and other presampling filters, there is no aliasing. In the following, we explain the transformation process step by step.

B. Inverting the HQ Operations
We attempt to invert as much as we can the operations of the HQ system, presented in Section 2 in reversed order (excluding treatment of noise, for the moment). We start by inverting the internal signal processing of the HQ system. This involves (beside quantization noise) two linear scale operations, ADC and DRC [Eqs. (16) and (17), respectively]. The signal processing of the HQ system is easy to invert once the parameters of the operations are known, resulting in the estimated detector readout $\hat{V}^{\text{HQ}}(m,n)$. Note that $\hat{V}^{\text{HQ}}(m,n)$ is not an analog signal as is $V^{\text{HQ}}(m,n)$. We just rescale the values.

We then invert the detector operations to estimate the photon irradiance (per sample). Inverting the readout operation is simply a multiplicative scale:

$$\hat{N}_e^{\text{HQ,samp}}(m,n) = C^{\text{HQ}}\hat{V}^{\text{HQ}}(m,n)/q \ [e]. \qquad (25)$$

As discussed above, we do not invert the noise addition, yet we calculate $\nu_{\text{noise}}^{\text{HQ}}$ from Eqs. (13) and (15) based on $\hat{N}_{\text{e}}^{\text{HQ,samp}}$ and continue with the inversion. Moreover, we do not invert the sampling operation. We leave the signal sampled as the HQ system has sampled it.

Next, we invert the PSF of the detector. There are various deconvolution methods. For instance, using Fourier analysis, the electron generation rate is therefore

$$\hat{N}_{\text{e}}^{\text{HQ}} = \frac{1}{t_{\text{int}}^{\text{HQ}}} \mathcal{F}^{-1}\left[\frac{\mathcal{F}\hat{N}_{\text{e}}^{\text{HQ,samp}}}{\mathcal{F}\hat{h}_{\text{det}}^{\text{HQ}}} \right], \qquad (26)$$

where $\mathcal{F}$ represents the discrete Fourier transform (recall that we null the frequency components of $\hat{N}_{\text{e}}^{\text{HQ}}$ that are outside the passband of the LQ system). Subsequently, the detector cross-talk blur is inverted:

$$\hat{N}_e^{\text{HQ,gen}} = \mathcal{F}^{-1}\left[\frac{\mathcal{F}\hat{N}_{\text{e}}^{\text{HQ}}}{\mathcal{F}\hat{h}_{\text{xtk}}^{\text{HQ}}} \right]. \qquad (27)$$

These calculations exploit the sampling done in Eq. (21). which lead to discrete PSFs $\hat{h}$.

Finally, we estimate an inversion of the photodetection process [Eq. (6)]. At this stage, note that the spectral in-

formation is lost. However, in the systems we use, the spectral quantum efficiency $\eta_Q(\lambda)$ is rather uniform at the relevant spectral region. Hence, the average quantum efficiency, $\bar{\eta}_Q^{\text{HQ}}$ is used. Following Eq. (6), the photon irradiance on the detector is estimated as

$$\hat{E}_{\text{det}}^{\text{HQ}}(m,n) = \frac{\hat{N}_{\text{e}}^{\text{HQ,gen}}(m,n) - i_{\text{dark}}^{\text{HQ}}/q}{\bar{\eta}_Q^{\text{HQ}}} \left[\frac{\text{photon}}{\text{s}} \right]. \qquad (28)$$

The last operations to be inverted axe those of the optical assembly (Subsection 2.A). First, we compensate for the extra radiance originating from system body temperature. The radiance contributions are calculated using the equations in Appendix A, integrated over each sample area A_{d}^{HQ} and over the spectral response $\eta_Q^{\text{HQ}}(\lambda)$. Therefore, instead of the spectral radiance contributions $E_\lambda(x,y)$, we calculate

$$\hat{E}(m,n) = A_{\text{d}}^{\text{HQ}}\int_\lambda \eta_Q^{\text{HQ}}(\lambda)E_\lambda(x,y)\text{d}\lambda \left[\frac{\text{photon}}{\text{s}} \right]. \qquad (29)$$

Thus, the estimated blurred scene contribution is

$$\hat{E}_{\text{blur}}^{\text{HQ}}(m,n) = \hat{E}_{\text{det}}^{\text{HQ}}(m,n) - \hat{E}_{\text{opt}}^{\text{HQ}} - \hat{E}_{\text{rfl}}^{\text{HQ}} - \hat{E}_{\text{body}}^{\text{HQ}}. \qquad (30)$$

Next, the optical blur [Eq. (2)] is inverted in the Fourier plane:

$$\hat{E}_{\text{proj}}^{\text{HQ}} = \mathcal{F}^{-1}\frac{\mathcal{F}\hat{E}_{\text{blur}}^{\text{HQ}}}{\mathcal{F}\hat{h}_{\text{opt}}^{\text{HQ}}}. \qquad (31)$$

Finally, we invert the scene projection operator [Eq. (1)] to estimate the scene radiance:

$$\hat{L}^{\text{scn}}(m,n) = \frac{\hat{E}_{\text{proj}}^{\text{HQ}}(m,n)}{\Omega_{\text{opt}}^{\text{HQ}}\tau_{\text{opt}}\tau_{\text{F}}} \left[\frac{\text{photon}}{\text{sr}\cdot\text{s}} \right]. \qquad (32)$$

C. Applying the LQ Operators
In Subsection 3.B we have inverted the HQ operations (excluding noise and sampling) to estimate the scene radiance $\hat{L}^{\text{scn}}(m,n)$. This is the input to the next phase, where we apply the LQ operations to estimate the desired output $\hat{I}_{\text{out}}^{LQ}$. These operations involve the operations discussed in Section 2 with the modifications discussed in Subsection 3.A.

Specifically, we add noise to the simulated image as given in Eq. (13). The noise STD is calculated as described in Eqs. (22) and (23).

4. EXPERIMENTS

A. Experimental Setup
To demonstrate the approach, we use images taken by a disposable thermal seeker, which is part of a guided missile. This is the LQ system. The HQ images are acquired by an HQ thermal camera. The LQ system uses a 1D column array detector of 120 elements $(30\ \mu\text{m}\times40\ \mu\text{m}$ each), which scans the scene horizontally to give a field of view of $2.3°\times1.7°$. The integration time $t_{\text{int}}^{\text{LQ}}$ is set by the scanning to 64 μs. The HQ system uses a 2D matrix detector with 240×320 elements $(30\ \mu\text{m}\times30\ \mu\text{m}$ each) giv-

ing a field of view of $2° \times 1.5°$. Here, t_{int}^{HQ} is set adaptively by the scene. The spectral range covered by both systems is $3-5$ μm.

We received images (kindly supplied by the manufacturer, Rafael Ltd.) of different scenes grabbed concur-

rently using the HQ system and the LQ system, I_{out}^{HQ} and I_{out}^{LQ} respectively. In addition, we received parameters that were grabbed with the HQ images such as t_{int}, T_{body}, and DRC parameters (g and o).

In order to implement our simulation, we need to input the imaging parameters of the HQ system and the LQ
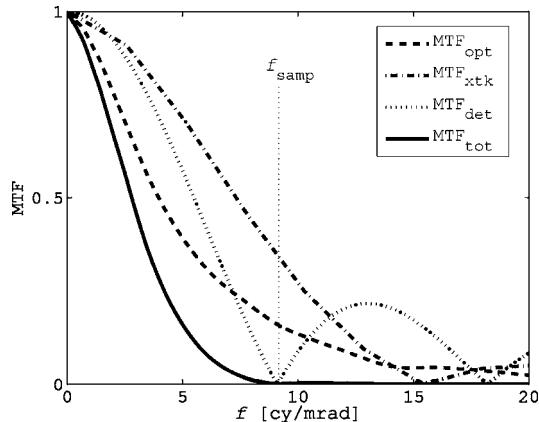
Fig. 8.   Presampling MTFs of the HQ system. The sampling frequency f_{samp} is marked on the plot.
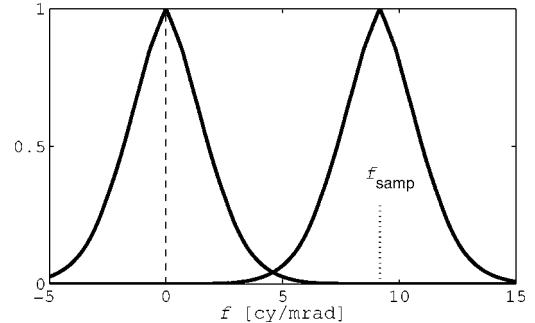
Fig. 9.   Spectrum associated with the entire HQ imaging system, before sampling. As a result of sampling it is replicated. There is very little energy in the overlap between the replicates: a very small portion of the spectrum is aliased.
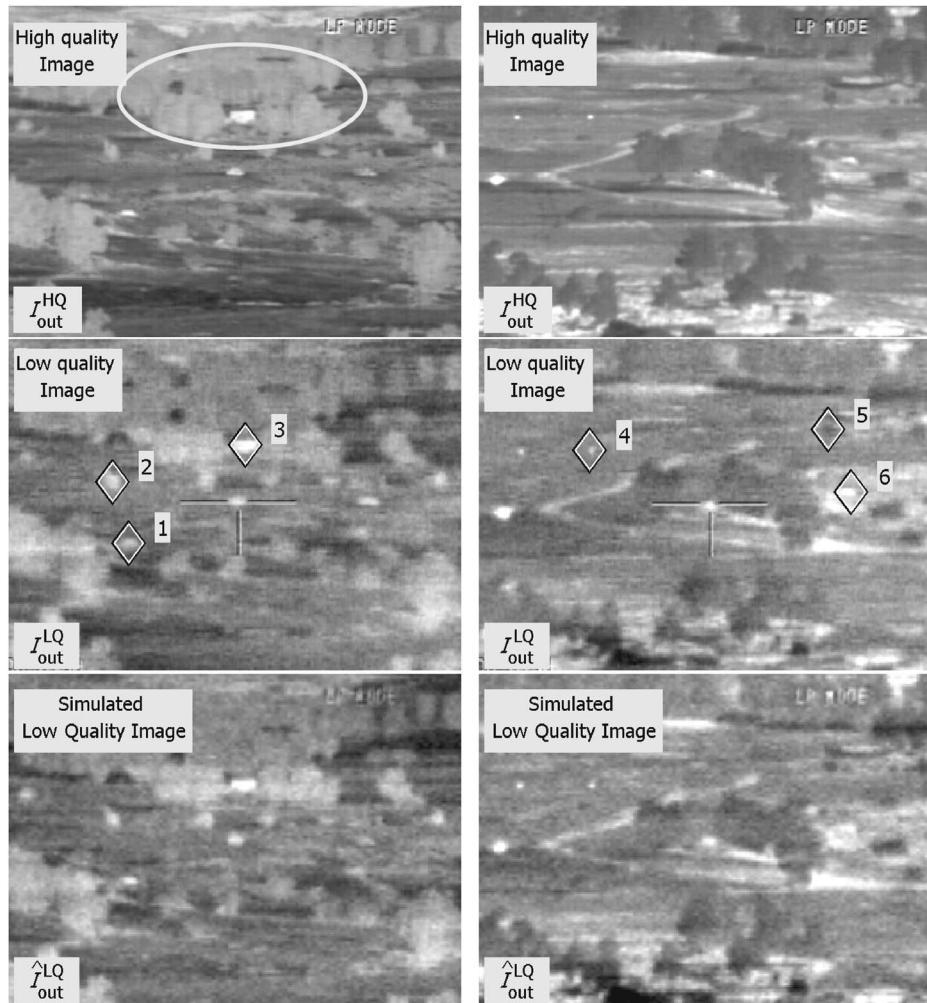
Fig. 10.   Experimental results. Each column includes a set from a distinct experiment. There is a significant difference between I_{out}^{HQ} and I_{out}^{LQ}. Nevertheless, $\hat{I}_{out}^{LQ}$ appears very similar to I_{out}^{LQ}. Targets are labeled for a lock-on grade comparison. Pay attention to the trees in the ellipse. In I_{out}^{HQ} they all have the same gray level, but in I_{out}^{LQ} they have different gray levels in different zones, an effect caused by the median subtraction. In $\hat{I}_{out}^{LQ}$ we see the same effect.
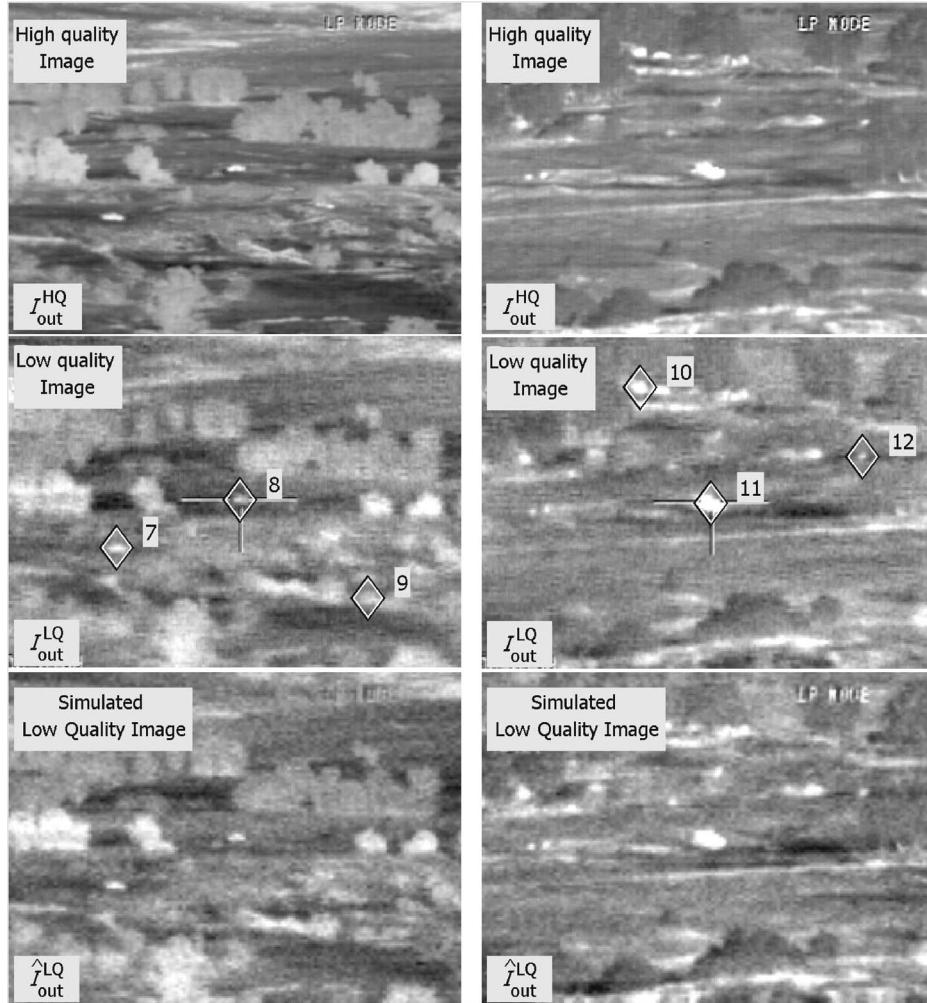
Fig. 11.   Additional experimental results. Each column includes a set from a distinct experiment. In all sets, there is a significant appearance difference between $I_{\text{out}}^{\text{HQ}}$ and $I_{\text{out}}^{\text{LQ}}$. Nevertheless, $\hat{I}_{\text{out}}^{\text{LQ}}$ appears very similar to $I_{\text{out}}^{\text{LQ}}$. Targets are labeled for a lock-on grade comparison.

system. The basic parameters (f_{opt}, A_{opt}, A_d, etc.) were given to us by the manufacturer, as well as the PSFs, which were measured by the manufacturer. Figure 7 plots the horizontal PSFs of the systems. It also indicates the horizontal pitch of the detector (30 μm). The optical PSF h_{opt} was estimated based on laboratory measurement of images of a narrow slit target in various directions. The measurement was performed by a small scanning detector. The cross-talk PSF h_{xtk} was measured in the following way: The detector was irradiated in a subpixel area (using a laser beam and a mask). The irradiated area was moved across the detector, and h_{xtk} was calculated from the readout. Figure 8 plots the presampling modulation transfer functions (MTFs) of the HQ system. The sampling frequency, f_{samp}, is marked on the plot. Figure 9 plots the spectrum associated with the entire HQ imaging system, before sampling. As a result of sampling, it is replicated. There is very little energy in the overlap between the replicates: a very small portion of the spectrum is aliased. Hence, aliasing is minor in this case.

To measure the noise STD, ν_{noise}, images were taken of a uniform target in the temperature of the scene. The temporal noise was then the temporal STD per pixel. Spa-

tial noise was estimated as the spatial STD of the image of the target, calculated after temporal frame averaging had eliminated the temporal noise. For a scene at room temperature the temporal noise STD, $\sqrt{\nu_{\text{shot}}^2 + \nu_{\text{excess}}^2}$, was equivalent to $\approx 1500e$ for the LQ system and $\approx 1350e$ for the HQ system. The spatial noise STD, ν_{RNU}, was equivalent to $\approx 800e$ for the HQ system. For the LQ system, ν_{RNU} could be neglected because of the median subtraction.

Using the parameters given to us, we implemented the cross-sensor transformation on the HQ images $I_{\text{out}}^{\text{HQ}}$, as explained in Section 3, using MATLAB to simulate the LQ images $\hat{I}_{\text{out}}^{\text{LQ}}$. In the following we compare the $I_{\text{out}}^{\text{HQ}}$ images with the grabbed LQ images $I_{\text{out}}^{\text{LQ}}$.

B. Visual Comparison
We first present a visual comparison. The images $I_{\text{out}}^{\text{HQ}}$, $I_{\text{out}}^{\text{LQ}}$, and $\hat{I}_{\text{out}}^{\text{LQ}}$ are shown in Fig. 10. There is a significant difference between $I_{\text{out}}^{\text{HQ}}$ and $I_{\text{out}}^{\text{LQ}}$, due to the difference between the systems. Nevertheless, $\hat{I}_{\text{out}}^{\text{LQ}}$ appears very similar to $I_{\text{out}}^{\text{LQ}}$. In particular, note the area surrounded by an ellipse. In $I_{\text{out}}^{\text{HQ}}$, one can see a building among trees, and all the trees have the same gray level. On the other hand, in

$I_{\text{out}}^{\text{LQ}}$ the trees are smeared, and they have different gray levels in different zones. The latter effect is caused by the median subtraction, which affects each row differently. The image $\hat{I}_{\text{out}}^{\text{LQ}}$ has the same effects. Additional sets are shown in Fig. 11, with similar behavior.

C. Quantitative Measure

We would like to compare the results quantitatively. To do this, we need a quantitative measure. We opted for a measure that is based on a computer-vision assignment, which is what the LQ system we use is designed for. Consider the task of lock-on by a tracker. Here, lock-on performance yields a quantitative measure for the computer vision task of *this system*. We need to demonstrate a performance similarity between $I_{\text{out}}^{\text{LQ}}$ and $\hat{I}_{\text{out}}^{\text{LQ}}$. This emulates the scenario of operation: Before activating the disposable system, there is a need to estimate the lock-on performance based on a live sample of the scene.

The lock-on quality is a function of variance and noise in a specified area around a target. Therefore, to calculate the lock-on grade for a target, a 15×15 pixel window (I_{win}) is set around the target. In this window the lock-on improves as the target contrast increases relative to the image noise. Thus, let the lock-on grade be

$$G_{\text{lock}} = \text{STD}(I_{\text{win}})/\nu_{\text{noise}}, \tag{33}$$

where ν_{noise} is the noise STD. For $\hat{I}_{\text{out}}^{\text{LQ}}$, the noise STD ν_{noise} is estimated as described in Subsection 3.A.3. We have calculated G_{lock} for different targets, which are labeled in Figs. 10 and 11. The lock-on grades on these targets are summarized in Table 1, estimated by nearby pixels around the target. Indeed, G_{lock} is significantly higher in $I_{\text{out}}^{\text{HQ}}$ than in $I_{\text{out}}^{\text{LQ}}$. On the other hand, G_{lock} in each target in $\hat{I}_{\text{out}}^{\text{LQ}}$ is about the same as in $I_{\text{out}}^{\text{LQ}}$. We thus conclude that in

Table 1. Lock-on Grade in Different Targets in $I_{\text{out}}^{\text{HQ}}$, $I_{\text{out}}^{\text{LQ}}$, and $\hat{I}_{\text{out}}^{\text{LQ}}$[a]

	G_{lock}		
Target	$I_{\text{out}}^{\text{HQ}}$	$I_{\text{out}}^{\text{LQ}}$	$\hat{I}_{\text{out}}^{\text{LQ}}$
1	19.4	4.5	4.3
2	14.1	4.1	4.2
3	28.7	6.1	5.9
4	11.1	2.0	2.1
5	11.9	3.2	3.1
6	17.2	4.0	3.8
7	21.9	5.8	5.6
8	21.9	3.7	3.9
9	9.6	4.2	4.0
10	18.0	9.6	9.8
11	30.0	10.9	10.7
12	15.8	3.9	4.1

[a]The grades in $I_{\text{out}}^{\text{HQ}}$ are an order of magnitude higher than those in $I_{\text{out}}^{\text{LQ}}$. Nevertheless, the grades in $\hat{I}_{\text{out}}^{\text{LQ}}$ are similar to those in $I_{\text{out}}^{\text{LQ}}$.
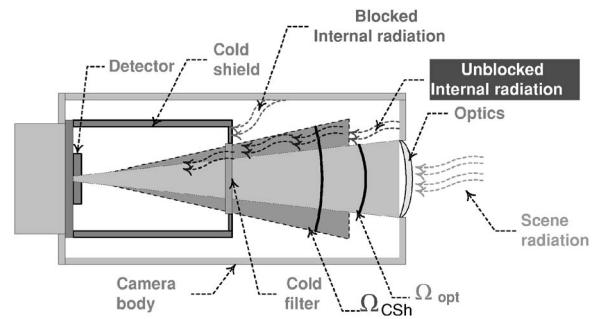
Fig. 12. Cold shield efficiency. The lighter cone shows the solid angle set by the optics Ω_{opt}, while the darker cone shows the solid angle set by the cold shield aperture Ω_{CSh}. When $\Omega_{\text{CSh}} > \Omega_{\text{opt}}$, some unwanted internal radiation reaches the detector.

these examples, simulating $\hat{I}_{\text{out}}^{\text{LQ}}$ based on real HQ image data leads to a good forecast of the performance of the LQ system in this computer-vision task.

5. DISCUSSION

We believe that our model is robust, since it is based on the physical parameters of the systems. It has two advantages over the mentioned example-based methods. First, it can deal with arbitrary transformations once the imaging process is known. Moreover, it is simple, fast, and noniterative, in contrast to current example-based methods. The latter deal with limited transformations, are iterative, and consume memory and processing resources.

We believe that this approach can be further extended to video sequences. There, temporal effects should be considered. For example, the noise model should handle spatial noise and temporal noise explicitly. In addition, temporal blur operators should be modeled. Furthermore, it would be interesting to apply these principles to other image modalities such as visible-light and medical environments.

APPENDIX A: AMBIENT RADIANCE CONTRIBUTIONS

In addition to the scene in view, a system may sense unrelated radiance components. These components come from the camera body and from the optics. Figure 3 depicts the different radiance contributions to the photon irradiance on the detector. In thermal imaging the nonscene contributions are significant. In the following, we explain the processes that each of these contributions undergoes. Since the radiation passes through a cold filter, all contributions are multiplied by the transmittance of this cold filter, $\tau_{\text{F}}(\lambda)$.

The first component is *direct* internal radiance. The camera body is hot; thus it radiates. Most of this radiation is blocked by a cold shield. However, as depicted in Fig. 12, part of this radiation may directly reach the detector. This occurs when the solid angle set by the cold shield aperture (Ω_{CSh}) is larger than the solid angle set by the optics (Ω_{opt}). This imperfection is expressed by the cold shield efficiency, $\eta_{\text{CSh}} = \Omega_{\text{opt}}/\Omega_{\text{CSh}}$. If the emissivity of the camera body is $\epsilon_{\text{b}}(\lambda)$, then the contribution of the direct radiation to the photon irradiance at the detector is

$$E_\lambda^{\text{body}} = \Omega_{\text{opt}}\left(1/\eta_{\text{CSh}} - 1\right)L_\lambda(T_b)\epsilon_b(\lambda)\tau_F(\lambda).$$

The second component is *reflected* internal radiance, which is part of camera body radiance, reflected by the optics toward the detector. It is given by

$$E_\lambda^{\text{rfl}} = L_\lambda(T_b)\epsilon_b(\lambda)\tau_F(\lambda)\int\int_{\Omega_{\text{opt}}}\rho_{\text{opt}}(\lambda,\theta,\phi)\mathrm{d}\theta\mathrm{d}\phi. \quad (34)$$

Here $\rho_{\text{opt}}(\lambda,\theta,\phi)$ is the reflectance of the optics toward the detector, where θ and ϕ are the angles of incidence at the optical aperture.

The third component is *optics* self-radiance: The optical components radiate. This radiance depends on the temperature of the optical components, which usually equals the camera body temperature. It also depends on the emissivity of the optics, $\epsilon_{\text{opt}}(\lambda)$. The photons are gathered through Ω_{opt}, thus contributing E_λ^{opt} $=\Omega_{\text{ope}}L_\lambda(T_b)\epsilon_{\text{opt}}(\lambda)\tau_F(\lambda)$ to the photon irradiance at the detector. Note that T_b may change during operation. In particular, it may increase by aerodynamic heating in high-speed missiles. Hence, image prediction may need to account for that. In our example, the lock-on task is performed when the missile is still stationary, and this is the task that should be predicted.

ACKNOWLEDGMENTS

REFERENCES

1. J. F. Bronskill, J. S. A. Hepburn, and W. K. Au, "A knowledge-based approach to the detection, tracking and classification of target formations in infrared image sequences," in *Proceedings of the IEEE Computer Society Conference Computer Vision and Pattern Recognition* (IEEE Computer Society, 1989), pp. 153–158.
2. L. E. Hoff, J. R. Evans, and L. E. Bunny, "Detection of targets in terrain clutter by using multi-spectral infrared image processing," Proc. SPIE **1481**, 98–109 (1991).
3. D. Sheffer and D. Ingman, "The informational difference concept in analayzing target recognition issues," J. Opt. Soc. Am. A **14**, 1431–1438 (1997).
4. R. G. Driggers, R. Vollmerhausen, and T. Edwards, "The target identification performance of infrared imager models as a function of blur and sampling," Proc. SPIE **3701**, 26–34 (1999).
5. J. Johnson, "Analysis of image forming systems," in *Proceedings of the Image Intensifier Symposium* (Warfare Electrical Engineering Department, U.S. Army Research and Development Laboratories, 1958), pp. 249–273.
6. L. Scott and J. D'Agostino, "NVEOD FLIR92 thermal imaging systems performance model," Proc. SPIE **1689**, 194–203 (1992).
7. W. Wittenstein, "Thermal range model TRM3," Proc. SPIE **3436**, 413–424 (1998).
8. O. E. Toler and D. S. Grey, "Simulation model for infrared imaging systems," Proc. SPIE **226**, 121–128 (1980).
9. M. Wegener and R. Drake, "High-fidelty synthetic IR imaging model," Proc. SPIE **4027**, 323–328 (2000).
10. C. Wigren, "Model of image generation in optronic sensor systems (IGOSS)," Proc. SPIE **3377**, 89–96 (1998).
11. W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE Computer Society, 1999), Vol. 2, pp. 1182–1189.
12. D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis." *Proceedings of SIGGRAPH 1995* (www.siggraph.org, 1995), pp. 229–238.
13. L. Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in *Proceedings of SIGGRAPH 2000* (www.siggraph.org, 2000), pp. 479–488.
14. P. Bao and M. Xiaohu, "Texturing and line art rendering using patch-based image analogies," in *Proceedings of the IEEE Workshop on Multimedia Signal Processing* (IEEE, 2002), pp. 142–148.
15. I. Drori, D. Cohen-Or, and H. Yeshurun, "Example-based style synthesis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, 2003), Vol. 2, pp. 143–150.
16. A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image analogies," in *Proceedings of SIGGRAPH 2001* (www.siggraph.org, 2001), pp. 327–340.
17. C. Oh, N. Nandhakurnar, and J. K. Aggarwal, "Integrated modelling of thermal and visual image generation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, 1989), pp. 356–362.
18. I. Pavlidis, J. Levine, and P. Baukol, "Thermal imaging for anxiety detection," in *Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum* (IEEE Computer Society, 2000), pp. 104–109.
19. D. A. Socolinsky, L. B. Wolff, J. D. Neuheisel, and C. K. Eveland, "Illumination invariant face recognition using thermal infrared imagery," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, 2001), Vol. 1, pp. I–527–534.
20. G. C. Holst, *Electro-Optical Imaging Systems Performance*, 3rd ed. (SPIE, 2003).
21. R. D. Hudson, *Infrared System Engineering* (Wiley, 1969), Chaps. 1–2, 5–9, 16–19.
22. J. M. Lloyd, *Thermal Imaging Systems* (Plenum, 1975), Chaps. 1–3, 5–11.
23. J. M. Mooney, "Design consideration for IR staring-mode cameras," Proc. SPIE **683**, 35–50 (1986).
24. C. A. Klein, "Hot infrared domes: a case study," Proc. SPIE **1326**, 217–230 (1990).
25. O. Naveh, "Sensitivity of scanning and staring infrared seekers for air-to-air missiles," Proc. SPIE **3061**, 692–711 (1997).
26. M. D. Nelson, J. F. Johnson, and T. S. Lomheim, "General noise process in hybrid infrared focal plane arrays," Opt. Eng. **30**, 1682–1700 (1991).
27. J. T. Barnett, "Statistical analysis of median subtraction filtering with application to point target detection in infrared backgrounds," Proc. SPIE **1050**, 10–18 (1989).
28. S. Nong and Z. Tianxu, "Segmentation of FLIR images by target enhancement and image model," Proc. SPIE **3545**, 274–277 (1998).