

Task-specific information for imaging system analysis

Mark A. Neifeld,^{1,2} Amit Ashok,^{1,*} and Pawan K. Baheti¹

¹*Department of Electrical and Computer Engineering, 1230 East Speedway Boulevard, University of Arizona, Tucson, Arizona 85721, USA*

²*College of Optical Sciences, 1630 East University Boulevard, University of Arizona, Tucson, Arizona 85721, USA*

*Corresponding author: ashoka@ece.arizona.edu

Received January 31, 2007; accepted May 15, 2007;

posted May 29, 2007 (Doc. ID 79534); published September 26, 2007

Imagery is often used to accomplish some computational task. In such cases there are some aspects of the imagery that are relevant to the task and other aspects that are not. In order to quantify the task-specific quality of such imagery, we introduce the concept of task-specific information (TSI). A formal framework for the computation of TSI is described and is applied to three common tasks: target detection, classification, and localization. We demonstrate the utility of TSI as a metric for evaluating the performance of three imaging systems: ideal geometric, diffraction-limited, and projective. The TSI results obtained from the simulation study quantify the degradation in the task-specific performance with optical blur. We also demonstrate that projective imagers can provide higher TSI than conventional imagers at small signal-to-noise ratios. © 2007 Optical Society of America

OCIS codes: 110.3000, 110.4280, 200.3050.

1. INTRODUCTION

The information content of an image plays an important role in a wide array of applications, ranging from video compression to imaging system design [1–6]. However, the computation of image information content remains a challenging problem. The problem is made difficult by (a) the high dimensionality of useful images, (b) the complex/unknown correlation structure among image pixels, and (c) the lack of relevant probabilistic models. It is possible to approximate the information content of an image by using some simplifying assumptions. For example, Gaussian and Markovian models have both been used to estimate image information [4,5,7]. Transform domain techniques have also been studied (e.g., wavelet prior models) [8,9].

As natural images possess a high degree of redundancy, it is generally understood that the information content of a natural image is not simply the product of the number of pixels and the number of bits per pixel. A very intuitive upper bound on the information content of an image can be obtained from the file size that is generated by a lossless compression algorithm. Consider the 256×256 pixel image shown in Fig. 1(a). An uncompressed version of this image requires 8 bits/pixel, resulting in a file size of 524,288 bits, whereas a lossless compression algorithm yields a file size of only 299,600 bits. A tighter upper bound might be obtained from a lossy compression algorithm that yields a visually indistinguishable reconstruction. Figure 1(b) depicts a reconstruction obtained using JPEG2000 [10], which yields a compressed file size of 36,720 bits. We may conclude from the high quality of this reconstruction that bits discarded from Fig. 1(a) to obtain Fig. 1(b) were not important to visual quality.

Imagery is often used in support of a computational task (e.g., automated target recognition). For this reason we would like to pursue a simple extension to the result shown in Fig. 1(b) in which the task performance, instead of visual quality, is the relevant metric. In such a scenario we might expect there to be aspects of the imagery that are important to the task and other aspects that are not. For example, if our task is target detection, then the image shown in Fig. 1(c) may contain nearly as much information as do the images in Figs. 1(a) and 1(b). The file size required for the image in Fig. 1(c) is only 25,120 bits. Taking this process one step further, a compression algorithm that actually *performs* target (a tank in this case) detection would yield a compressed file size of only 1 bit to indicate either “target present” or “target absent.” The preceding discussion demonstrates that an image used for target detection will contain no more than 1 bit of *relevant* information. We will refer to this relevant information as task-specific information (TSI), and the remainder of this paper represents an effort to describe/quantify TSI as an analysis tool for several tasks and imaging systems of interest. What we describe here is a formal approach to the computation of TSI. Such a formalism is important primarily because it enables imager design and/or adaptation that strives to maximize the TSI content of measurements. This has two implications: (a) imager resources can be optimally allocated so that irrelevant information is not measured and thus task-specific performance is maximized, and/or (b) imager resources can be minimized subject to a TSI constraint, thus reducing imager complexity, cost, size, weight, etc. It is worth mentioning that as TSI is a Shannon information-theoretic measure it can be used to bound conventional task perfor-

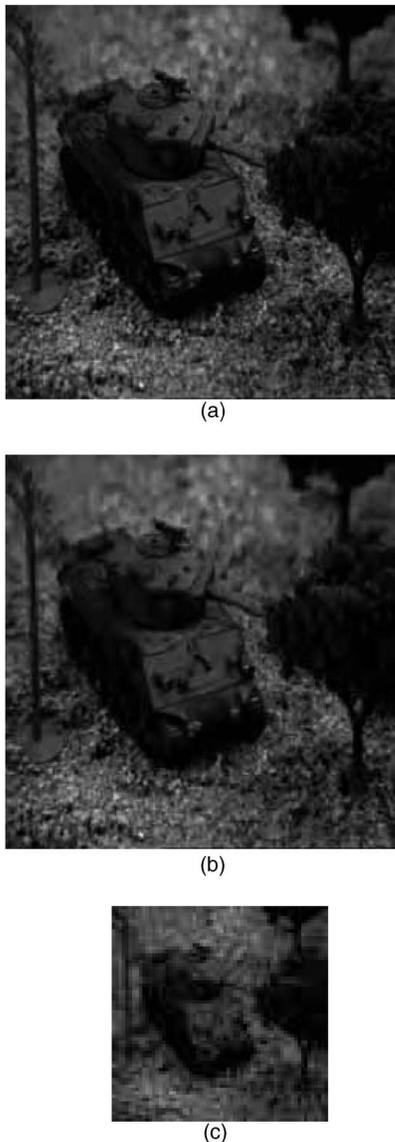


Fig. 1. (a) 256×256 image, (b) compressed version of the image in (a) using JPEG2000, and (c) 64×64 image obtained by rescaling the image in (a).

mance metrics, such as probability of error, via Fano’s inequality for a classification task [11].

Although a formal approach for quantifying the Shannon information in a task-specific way has not been previously reported, we do note important previous work concerning the use of task-based metrics for image quality assessment by Barrett *et al.* [12–15]. This previous work has focused primarily on ideal observer models and their application to various detection and estimation tasks.

The remainder of this paper is organized as follows: Section 2 introduces a formal framework for the definition of TSI and a method for its computation using conditional mean estimators. We consider three example tasks: target

detection, target classification, and joint detection/classification and localization. In Section 3 we apply the TSI framework to two simple imaging systems: an ideal geometric imager and a diffraction-limited imager for each of the three tasks. Section 4 extends the imaging model to projective imagers. The TSI framework is applied to the analysis of two projective imagers: a principal component (PC) projective imager and a matched-filter (MF) projective imager. Section 5 summarizes the TSI framework and draws the final conclusions.

2. TASK-SPECIFIC INFORMATION

We begin by considering the various components of an imaging system. A block diagram depicting these components is shown in Fig. 2. In this model, the scene Y provides the input to the imaging channel represented by the operator \mathcal{H} to yield $Z = \mathcal{H}(Y)$. The quantity Z is then corrupted by the noise operator \mathcal{N} to yield the noisy measurement $R = \mathcal{N}(Z)$. The model in Fig. 2 is made task specific via the incorporation of the virtual source and encoding blocks. The virtual source variable X represents the parameter of interest for a specific task. For example, a target detection task would utilize a binary-valued virtual source variable to indicate the presence ($X=1$) or absence ($X=0$) of the target. Note that this virtual source serves as a mechanism through which we can quantify the TSI embedded in a scene. The encoding operator \mathcal{C} uses X to generate the scene according to $Y = \mathcal{C}(X)$. In general, \mathcal{C} can be either deterministic or stochastic. In order to illustrate how \mathcal{C} generates a *scene*, let us consider the following two examples.

Our first example demonstrates the use of a deterministic encoding specified by the operator

$$C_{S1}(X) = \vec{V}_{target}X + \vec{V}_{bg}, \tag{1}$$

where C_{S1} is a deterministic operator, the virtual source X is a binary random variable, \vec{V}_{target} represents the target profile, and \vec{V}_{bg} is the background profile. Note that \vec{V}_{target} and \vec{V}_{bg} are vectors formed by unrastring a two-dimensional image into a column vector. Figures 3(a) and 3(b) show the encoder output for $X=1$ and $X=0$, respectively. The scene model defined by C_{S1} could be useful in a problem where the task is to detect the presence or absence of a known target at a known position in a known background.

Our second example demonstrates the use of a stochastic encoding specified by the operator

$$C_{S2}(X) = \vec{V}_{target}X + \vec{V}_{bg} + \vec{V}_{tree}\beta_1 + \vec{V}_{shrub}\beta_2, \tag{2}$$

where X , \vec{V}_{target} , and \vec{V}_{bg} are the same as in Eq. (1). Clutter components \vec{V}_{tree} and \vec{V}_{shrub} represent tree and shrub profiles, respectively, and are weighted by random variables β_1 and β_2 . Note that for a given value of X , C_{S2} will

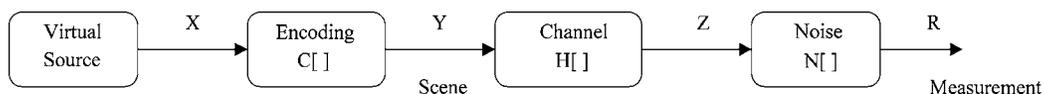


Fig. 2. Block diagram of an imaging chain.

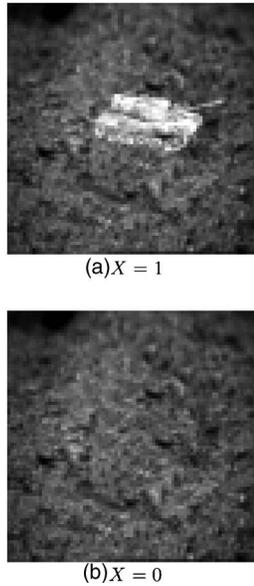


Fig. 3. Example scenes from the deterministic encoder.

depend on random variables β_1 and β_2 ; therefore, C_{S2} is a stochastic operator. Figures 4(a) and 4(b) show examples of scene realizations generated by this stochastic encoding operator.

As X is the only parameter of interest for a given task, it is important to note that the entropy of X defines the maximum TSI content of any image measurement. Other blocks in the imaging chain may add entropy to the image measurement R ; however, only the entropy of the virtual source X is relevant to the task. We may therefore define TSI as the Shannon mutual information $I(X;R)$ between the virtual source X and the image measurement R as follows [11]:

$$\text{TSI} \equiv I(X;R) = J(X) - J(X|R), \quad (3)$$

where $J(X) = -\mathbb{E}\{\log(pr(X))\}$ denotes the entropy of virtual source X , $J(X|R) = -\mathbb{E}\{\log(pr(X|R))\}$ denotes the entropy of X conditioned on the measurement R , $\mathbb{E}\{\cdot\}$ denotes statis-

tical expectation, $pr(\cdot)$ denotes the probability density function, and all the logarithms are taken to be base 2. Note that from this definition of TSI we have $I(X;R) \leq J(X)$, indicating that an image cannot contain more TSI than there is entropy in the variable representing the task. However, for most realistic imaging problems, computing TSI from Eq. (3) directly is intractable owing to the dimensionality and non-Gaussianity of R . Numerical approaches may also prove to be computationally prohibitive, even when using methods such as importance sampling, the Markov-chain Monte Carlo (MCMC) technique or the Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm [16–20].

Recently, Guo *et al.* [21] demonstrated a direct relationship between the minimum mean-square error (*mmse*) in estimating X from R , and the mutual information $I(X;R)$ for an additive Gaussian channel. Although the relation between estimation *mmse* and Fisher information has been known via Van Tree's inequality [22], Guo's result connects estimation *mmse* with the Shannon information for the first time. The result expresses *mmse* as a derivative of the mutual information $I(X;R)$ with respect to the signal-to-noise ratio. For a simple additive Gaussian noise channel we have

$$R = \sqrt{s}X + N, \quad (4)$$

where N is additive Gaussian noise with variance $\sigma^2=1$ and s is the signal-to-noise ratio. For this simple case we find that [21]

$$\frac{d}{ds}I(X;R) = \frac{1}{2}mmse = \frac{1}{2}\mathbb{E}[|X - \mathbb{E}(X|R)|^2], \quad (5)$$

where $\mathbb{E}(X|R)$ is the conditional mean estimator. This relation allows us to compute mutual information indirectly from *mmse* for an additive Gaussian channel without any restrictions on the distribution of the virtual source variable X . It is interesting to note that even though the source variable X is discrete valued, the conditional mean estimator is a continuous variable that does not necessarily take values in the range of the source variable X . For

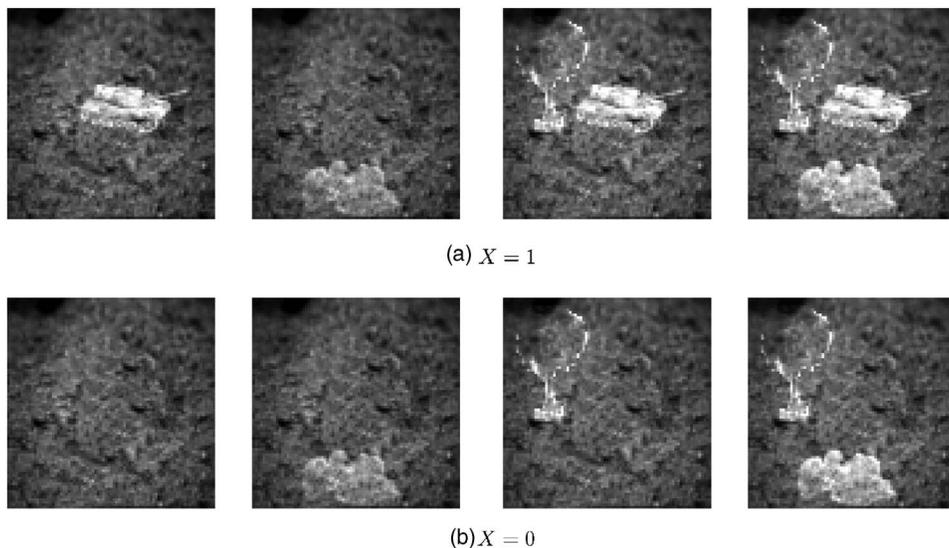


Fig. 4. Example scenes from the stochastic encoder.

example, when X is a binary variable(0/1) the conditional mean estimator will yield a real number between 0 and 1.

The scalar result in Eq. (5) has been extended to the linear vector Gaussian channels for which $\mathcal{H}[\vec{X}] = \mathbf{H}\vec{X}$, where \mathbf{H} denotes the matrix channel operator and \vec{X} is the vector channel input. The output of such a channel can be written as

$$\vec{R} = \sqrt{s}\mathbf{H}\vec{X} + \vec{N}, \quad (6)$$

where \vec{N} follows a multivariate Gaussian distribution with covariance $\Sigma_{\vec{N}}$. In this case, Guo's result becomes [23]

$$\frac{d}{ds}I(\vec{X};\vec{R}) = \frac{1}{2}\mathbb{E}[\|\mathbf{H}\vec{X} - \mathbb{E}[\mathbf{H}\vec{X}|\vec{R}]\|^2]. \quad (7)$$

The right-hand side of Eq. (7) is the *mmse* in estimating $\mathbf{H}\vec{X}$ rather than \vec{X} , and therefore we denote it by *mmse_H* throughout the rest of the paper to avoid confusion. For an arbitrary noise covariance $\Sigma_{\vec{N}}$, *mmse_H* can be computed using $\text{Tr}(\mathbf{H}^\dagger \Sigma_{\vec{N}}^{-1} \mathbf{H} \mathbf{E})$, where $\mathbf{E} = \mathbb{E}[(\vec{X} - \mathbb{E}[\vec{X}|\vec{R}]) (\vec{X} - \mathbb{E}[\vec{X}|\vec{R}])^T]$, \mathbf{H}^\dagger denotes the Hermitian conjugate of \mathbf{H} , and $\text{Tr}(\cdot)$ denotes the trace of the matrix. Therefore, the relationship between mutual information and *mmse_H* can be written as

$$\frac{d}{ds}I(\vec{X};\vec{R}) = \frac{1}{2}mmse_H = \frac{1}{2}\text{Tr}(\mathbf{H}^\dagger \Sigma_{\vec{N}}^{-1} \mathbf{H} \mathbf{E}). \quad (8)$$

These results have also been extended to the case for which the channel input is a random function of \vec{X} , denoted by $\vec{Y} = \mathbf{C}(\vec{X})$. The relation between $I(\vec{X};\vec{R})$ and *mmse_H* for a random function $\mathbf{C}(\vec{X})$ is slightly different from the previous expression in Eq. (8). Using the stochastic encoding model we have

$$\vec{R} = \sqrt{s}\mathbf{H}\mathbf{C}(\vec{X}) + \vec{N}. \quad (9)$$

In this case the relation between mutual information and *mmse* can be expressed as [23]

$$\frac{d}{ds}I(\vec{X};\vec{R}) = \frac{1}{2}mmse_H, \quad (10)$$

where

$$mmse_H = \text{Tr}(\mathbf{H}^\dagger \Sigma_{\vec{N}}^{-1} \mathbf{H} (\mathbf{E}_{\vec{Y}} - \mathbf{E}_{\vec{Y}|\vec{X}})),$$

$$\mathbf{E}_{\vec{Y}} = \mathbb{E}[(\vec{Y} - \mathbb{E}[\vec{Y}|\vec{R}])(\vec{Y} - \mathbb{E}[\vec{Y}|\vec{R}])^T],$$

$$\mathbf{E}_{\vec{Y}|\vec{X}} = \mathbb{E}[(\vec{Y} - \mathbb{E}[\vec{Y}|\vec{R},\vec{X}])(\vec{Y} - \mathbb{E}[\vec{Y}|\vec{R},\vec{X}])^T].$$

Next we consider the application of these results to an important class of imaging problems. We make the following assumptions about the general imaging chain model: (1) the channel operator \mathcal{H} is linear (discrete to discrete) and deterministic, (2) the encoding operator \mathcal{C} is linear and stochastic, and (3) the noise model \mathcal{N} is additive and Gaussian. We begin by developing some basic scene models for the tasks of detection and classification.

A. Detection with Deterministic Encoding

For pedagogical purposes we begin with a scalar channel and a deterministic encoding. Consider a simple task of detecting the presence or absence of a known scalar signal t in the presence of noise. The measurement R is given as

$$R = \sqrt{s}t \cdot X + N, \quad (11)$$

where X is the virtual source variable that determines the signal present or absent condition and N represents additive white Gaussian noise (AWGN) with variance $\sigma^2=1$. Note that here the encoding operator is deterministic and is defined as $\mathcal{C}_s(X) = t \cdot X$. For simplicity, we set $\mathbf{H}Y = Y$. Because X is a binary random variable with probability distribution $\text{Pr}(X=1) = p$ and $\text{Pr}(X=0) = 1-p$, we can assert

$$I(X;R) \leq J(X) \leq 1 \text{ bit}, \quad (12)$$

where entropy of X is $J(X) = -p \log(p) - (1-p) \log(1-p)$. Note that for this simple detection task the received signal R contains at most 1 bit of TSI. Therefore, the performance of any detection algorithm that operates on the measurement R is upper bounded by the TSI.

We compute the mutual information $I(X;R)$ using two methods. The direct method is based on the definition of mutual information given in Eq. (3), wherein differential entropies will be used owing to the continuous-valued nature of R . The conditional differential entropy $J(R|X)$ equals $J(N) = \frac{1}{2} \ln(2\pi e \sigma^2)$. Note that $J(R)$ is not straightforward to compute, as R follows a mixture of Gaussian distribution defined as

$$pr(R) = \frac{1}{\sqrt{2\pi\sigma^2}} \left(p \exp \left[-\frac{(R - \sqrt{s}t)^2}{2\sigma^2} \right] + (1-p) \times \exp \left[-\frac{R^2}{2\sigma^2} \right] \right). \quad (13)$$

We therefore resort to numerical integration to compute $J(R)$. Note that when R is a vector this approach quickly becomes computationally prohibitive as the dimensionality of R increases.

The alternative method for computing $I(X;R)$ exploits the relationship between *mmse* and mutual information as stated in Eq. (5), where $\mathbb{E}(X|R)$ is the conditional mean estimator, which can be expressed as

$$\mathbb{E}(X|R) = \left[1 + \frac{1-p}{p} \exp \left(\frac{\sqrt{s}t(\sqrt{s}t - 2R)}{2\sigma^2} \right) \right]^{-1}. \quad (14)$$

The mutual information is computed by numerically integrating *mmse* over a range of s . The *mmse* itself is estimated using the Monte Carlo and importance-sampling methods [16–19].

Figure 5(a) shows a plot of *mmse* versus s for $p = \frac{1}{2}$ and $t = 1$. As expected, the *mmse* reduces with increasing s . The mutual information computed from this *mmse* data is plotted in Fig. 5(b) versus s . The curve with circle symbol corresponds to the mutual information computed using the *mmse*-based method, and the curve with plus symbol corresponds to the mutual information computed using the direct method as per Eq. (3). As expected, these two methods yield the same result. Note that Guo's method of estimating TSI via *mmse* is significantly more computa-

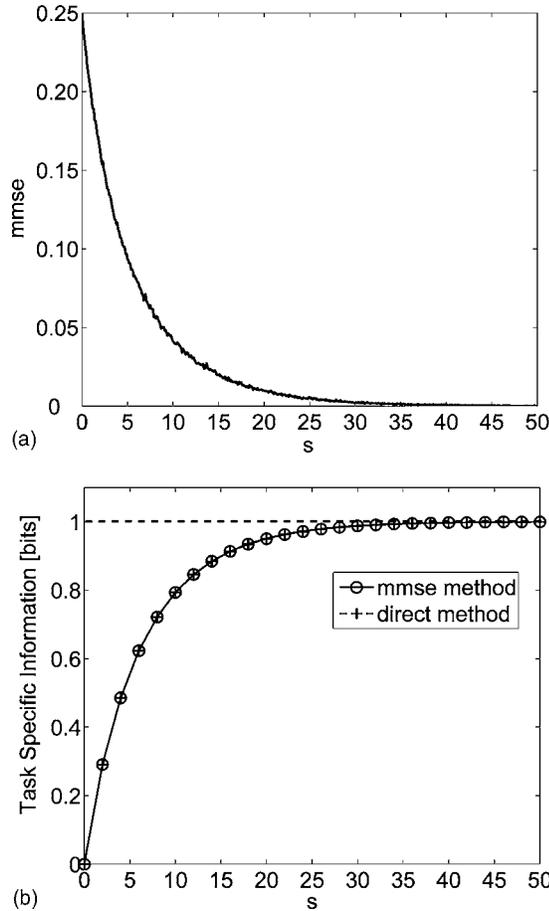


Fig. 5. Plot of (a) *mmse* and (b) TSI versus the signal-to-noise ratio for the scalar detection task.

tionally tractable for high-dimensional vector \vec{R} as compared with the direct method. Henceforth, all the TSI results reported herein will employ Guo's *mmse*-based method. Our pedagogical example considered a deterministic \mathcal{C} ; however, in any realistic scenario \mathcal{C} will be stochastic. Next we consider a detection task in which \mathcal{C} is stochastic, allowing for additional scene variability arising from random background and target realizations.

B. Detection with Stochastic Encoding

Let us consider a slightly more complex detection task where a known target is to be detected in the presence of noise and clutter. The target position is assumed to be variable and unknown, and hence for the detection task, the target position assumes the role of a nuisance parameter. Here we have considered only one nuisance parameter; however, more realistic scene models would utilize a multitude of nuisance parameters such as target orientation, location, magnification, etc. Our aim here is to demonstrate an application of the TSI framework, and the extension to additional nuisance parameters will be straightforward.

The imaging model for this task is constructed as

$$\vec{R} = \mathbf{H}\mathbf{C}_{\text{det}}(X) + \vec{N}, \quad (15)$$

where \mathbf{H} is the imaging channel matrix operator, \vec{N} is the zero-mean AWGN with covariance $\Sigma_{\vec{N}}$, and \mathbf{C}_{det} is the sto-

chastic encoding operator. The encoding operator \mathbf{C}_{det} is defined as

$$\mathbf{C}_{\text{det}}(X) = \sqrt{s}\mathbf{T}\vec{\rho}X + \sqrt{c}\mathbf{V}_c\vec{\beta}, \quad (16)$$

where \mathbf{T} is the target profile matrix in which each column is a target profile (lexicographically ordered into a one-dimensional vector) at a specific position in the scene. In general, when the scene is of dimension $M \times M$ pixels and there are P different possible target positions, the dimension of matrix \mathbf{T} is $M^2 \times P$. The column vector $\vec{\rho}$ is a random indicator vector and selects the target position for a given scene realization. Therefore, $\vec{\rho} \in \{\vec{c}_1, \vec{c}_2, \dots, \vec{c}_P\}$, where \vec{c}_i is a P -dimensional unit column vector with a 1 in the i th position and 0 in all remaining positions. Figure 6(a) illustrates the structure of \mathbf{T} and $\vec{\rho}$. Note that $\vec{\rho} = \vec{c}_2$ in Fig. 6(a), and therefore the output of $\mathbf{T}\vec{\rho}$ is the target profile at position 2. All positions are assumed to be equally probable; therefore $\Pr(\vec{\rho} = \vec{c}_i) = 1/P$ for $i = \{1, 2, \dots, P\}$. The virtual source variable X takes the value 1 or 0 (i.e., "target present" or "target absent") with probabilities p and $1-p$, respectively.

Here \mathbf{V}_c is the clutter profile matrix whose columns represent various clutter components such as tree, shrub, grass, etc. The dimension of \mathbf{V}_c is $M^2 \times K$, where K is the

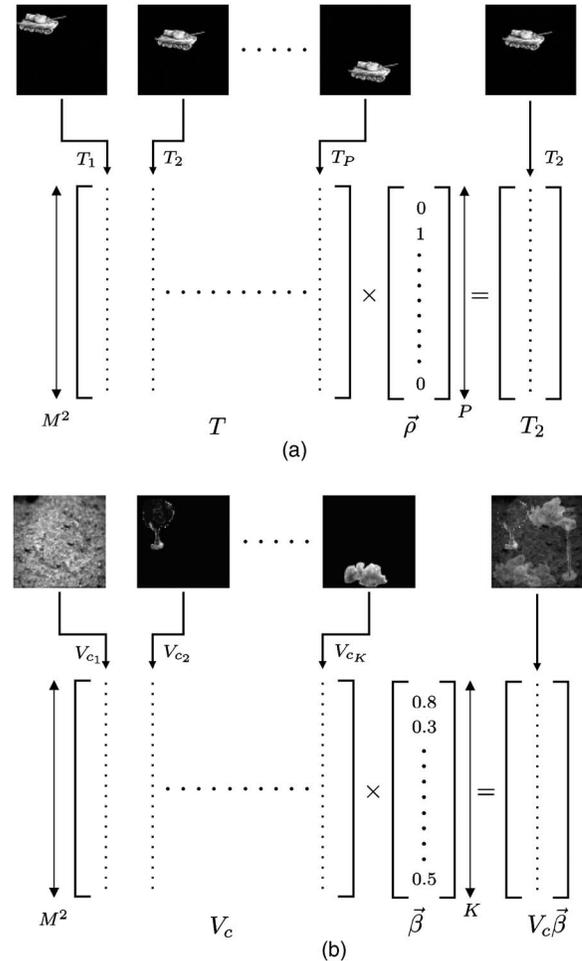


Fig. 6. Illustration of stochastic encoding \mathbf{C}_{det} : (a) target profile matrix \mathbf{T} and position vector $\vec{\rho}$ and (b) clutter profile matrix \mathbf{V}_c and mixing vector $\vec{\beta}$.

number of clutter components and $\vec{\beta}$ is the K -dimensional clutter mixing column vector, which determines the strength of various components that comprise the clutter. Note that $\vec{\beta}$ follows a multivariate Gaussian distribution with mean $\vec{\mu}_{\vec{\beta}}$ and covariance $\Sigma_{\vec{\beta}}$. Figure 6(b) shows individual clutter components arranged columnwise in the clutter profile matrix \mathbf{V}_c . The particular realization of clutter mixing vector $\vec{\beta}$ shown in Fig. 6(b) yields the clutter shown on the right-hand side.

The coefficient c in Eq. (16) denotes the clutter-to-noise ratio. Note that clutter and detector noise combine to form a multivariate Gaussian random vector $\vec{N}_c = \sqrt{c}\mathbf{H}\mathbf{V}_c\vec{\beta} + \vec{N}$ with mean $\vec{\mu}_{\vec{N}_c} = \sqrt{c}\mathbf{H}\mathbf{V}_c\vec{\mu}_{\vec{\beta}}$ and covariance $\Sigma_{\vec{N}_c} = \mathbf{H}\mathbf{V}_c\Sigma_{\vec{\beta}}\mathbf{V}_c^T\mathbf{H}^T + c\Sigma_{\vec{N}}$. Now we can rewrite the imaging model as

$$\vec{R} = \sqrt{s}\mathbf{H}\mathbf{T}\vec{\rho}X + \vec{N}_c. \quad (17)$$

The TSI for the detection task is the mutual information between the image measurement \vec{R} and the virtual source X . Since the encoding operator \mathbf{C}_{det} is a random function of the source X , we apply the result given in Eq. (10). Comparing Eq. (10) with the imaging model shown in Eq. (17), we note that the \vec{X} and \vec{Y} in Eq. (10) are equal to the virtual source X and $\mathbf{T}\vec{\rho}X$, respectively. The channel operator \mathcal{H} is substituted by \mathbf{H} and \vec{N} is replaced by \vec{N}_c in Eq. (10). The TSI and $mmse_H$ are therefore related as

$$\text{TSI} = I(X; \vec{R}) = \frac{1}{2} \int_0^s mmse_H(s') ds', \quad (18)$$

where

$$mmse_H(s) = \text{Tr}(\mathbf{H}^\dagger \Sigma_{\vec{N}_c}^{-1} \mathbf{H} (\mathbf{E}_{\vec{Y}} - \mathbf{E}_{\vec{Y}|X})), \quad (19)$$

$$\vec{Y} = \mathbf{T}\vec{\rho}X. \quad (20)$$

Explicit expressions for the estimators required for evaluating the expectations in Eq. (19) are derived in Appendix A.

C. Classification with Stochastic Encoding

We consider a simple two-class classification problem for which we label the two possible states of nature (i.e., tar-

gets) as H_1 and H_2 . The extension to more than two classes will be straightforward. The overall imaging model remains the same as in Eq. (15). The number of positions that each target can take remains unchanged. However, now \mathbf{T} has dimensions $M^2 \times 2P$ and is given by $\mathbf{T} = [\mathbf{T}_{H_1} \quad \mathbf{T}_{H_2}]$, where \mathbf{T}_{H_i} is the target profile matrix for class i . The structure of this composite target profile matrix \mathbf{T} is shown in Fig. 7. The virtual source variable is denoted by the vector \vec{X} and takes the values $[1, 0]^T$ or $[0, 1]^T$ to represent H_1 or H_2 , respectively. The prior probabilities for H_1 and H_2 are p and $1-p$, respectively. The vector $\vec{\rho}$ from the detection problem becomes a matrix ρ of dimension $2P \times 2$ and is defined as

$$\rho = \begin{bmatrix} \vec{\rho}_H & \mathbf{0} \\ \mathbf{0} & \vec{\rho}_H \end{bmatrix}, \quad (21)$$

where $\vec{\rho}_H \in \{\vec{c}_1, \vec{c}_2, \dots, \vec{c}_P\}$ and $\mathbf{0}$ is an all-zero P -dimensional column vector. Once again we assume all positions to be equally probable; therefore $\Pr(\vec{\rho}_H = \vec{c}_i) = 1/P$ for $i = \{1, 2, \dots, P\}$.

Consider an example that illustrates how the term $\mathbf{T}\rho\vec{X}$ enables selection of a target from either H_1 or H_2 at one of P positions. In order to generate a target from H_1 at the m th position in the scene, $\vec{\rho}_H = \vec{c}_m$ and $\vec{X} = [1, 0]^T$. The product of $\mathbf{T}\rho$ will produce an $M^2 \times 2$ matrix whose first column is equal to the H_1 profile at position m and whose second column is equal to the H_2 profile at the same position. This resulting matrix, when multiplied by $\vec{X} = [1, 0]^T$, will select the H_1 profile. Similarly, in order to choose a target from H_2 at the m th position, $\vec{\rho}_H = \vec{c}_m$ and $\vec{X} = [0, 1]^T$. Note that $\vec{\rho}_H = \vec{c}_2$ in Fig. 7 and therefore selects the second position for H_1 and H_2 .

The imaging model presented for the detection problem in Eq. (17) and the corresponding TSI defined in Eq. (18) require minor modifications to remain valid for the classification problem. Specifically, we require the virtual source variable to become a vector quantity \vec{X} and the dimensions of \mathbf{T} and ρ to be adjusted accordingly, as noted above. Note that despite the increase in dimensionality, the binary source vector \vec{X} results in the upper bound $\text{TSI} \leq 1$ bit for the two-class classification problem.

The two-class model for target classification can easily be extended to the case of joint detection and classifica-

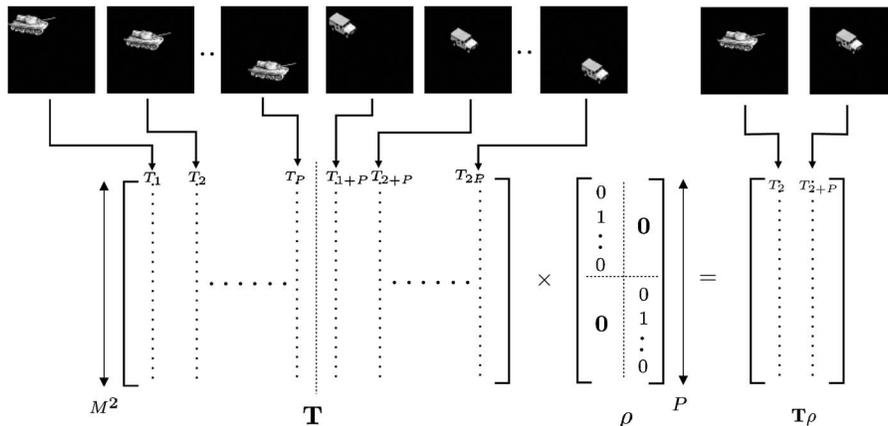


Fig. 7. Structure of \mathbf{T} and ρ matrices for the two-class problem.

tion. The simple extension involves introducing a third class corresponding to the null hypothesis and can be accommodated by allowing \vec{X} to also take the value $[0, 0]^T$ with some probability p_0 . The TSI upper bound in this case becomes $J(\vec{X}) = -p_0 \log(p_0) - p_1 \log(p_1) - p_2 \log(p_2) \leq 1.6$ bits for $p_0 = p_1 = p_2 = 1/3$. This important extension to joint detection and classification is pursued further in the next section, where we also consider the simultaneous estimation of an unknown target parameter.

D. Joint Detection/Classification and Localization

We begin with a discussion of the localization task. Later in this section we combine the encoding model for localization with the models for detection and classification described in Subsections 2.B and 2.C. Consider the problem of localizing a target (known to be present) in one of Q regions in a scene. The example shown in Fig. 8 depicts a case in which there are four regions ($Q=4$). Note that for this problem, the specific target location within a region is unimportant and is therefore treated as a nuisance parameter. We allow P_i possible target locations within the i th region such that $\sum_{i=1}^Q P_i = P$, where P is the total number of possible target locations in the scene. The noise and clutter models remain unchanged from Subsections 2.B and 2.C so that the task-specific imaging model for localization can be written as

$$\vec{R} = \sqrt{s} \mathbf{HT} \Lambda(X) \vec{\rho} + \vec{N}_c, \tag{22}$$

where we have simply inserted the localization matrix $\Lambda(X)$ into the channel model in Eq. (17). As defined earlier, the columns of \mathbf{T} correspond to the target profiles at all possible positions. For the sake of convenience, we rearrange the columns of \mathbf{T} such that the first P_1 columns represent the target profiles at the P_1 positions in region 1, the next P_2 columns correspond to region 2, and so on. The virtual source variable X is now a Q -ary variable, i.e., $X \in \{1, 2, \dots, Q\}$ representing one of the Q regions where the target is present and $\Lambda(X)$ acts as the localization matrix and selects all target profiles in the region specified by the source X . For the case $X=i$, $\Lambda(X=i)$ is of dimension $P \times P_i$ and is given by

$$\Lambda(X=i) = \begin{bmatrix} [\mathbf{0}]_{P_1 \times P_i} \\ \vdots \\ [\mathbf{0}]_{P_{i-1} \times P_i} \\ [\mathbf{I}]_{P_i \times P_i} \\ [\mathbf{0}]_{P_{i+1} \times P_i} \\ \vdots \\ [\mathbf{0}]_{P_Q \times P_i} \end{bmatrix}.$$

For $X=i$, $\vec{\rho}$ is a P_i -dimensional random indicator vector that selects one of the P_i target profiles resulting from $\mathbf{T} \Lambda(X=i)$. Therefore, $\vec{\rho} \in \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_{P_i}\}$, where \vec{e}_k is a P_i -dimensional unit column vector with a 1 in the k th position and 0 in all remaining positions. All positions within each region are considered to be equally probable; therefore, $\Pr(\vec{\rho} = \vec{e}_k) = [\Pr(X=i)]/P_i$, where $\Pr(X=i)$ is the probability of the target being located in region i and $k = \{1, 2, \dots, P_i\}$. Figure 8 illustrates the structure of \mathbf{T} and $\Lambda(X)$ using an example where $X=2$. In this example, P positions are equally distributed among the four regions, i.e., $P_i = P/4$ for $i = \{1, 2, 3, 4\}$. Observe that $\mathbf{T} \Lambda(X)$ selects all the target positions in region 2 and the postmultiplication of this matrix with $\vec{\rho} = \vec{e}_k$ results in the target at the k th position of region 2. Recall that the localization task is concerned only with estimating the region in which the target is present and the exact position within the region is treated as a nuisance parameter. Therefore, the upper bound on TSI in this case becomes $J(X) = -\sum_{q=1}^Q \Pr(X=q) \log \Pr(X=q) \leq [\log Q]$ bits.

We now combine the encoding model for localization, defined in Eq. (22), with the detection and classification models described in the previous section. For the joint detection/localization task, we are interested in detecting the presence of a target and, if present, localizing it in one of Q regions. The imaging model from Eq. (22) becomes

$$\vec{R} = \sqrt{s} \mathbf{HT} \Lambda(X) \vec{\rho} \alpha + \vec{N}_c, \tag{23}$$

where α is a binary variable indicating the presence or absence of the target. Therefore, the virtual source in this case is a $(Q+1)$ -ary variable and is defined as $X' \in \{X, 0\}$ so that when $\alpha=0$, $X'=0$ and when $\alpha=1$, $X'=X$. Comparing Eq. (10) with the imaging model shown in Eq. (23), we note that the \vec{X} and \vec{Y} in Eq. (10) are equal to the virtual

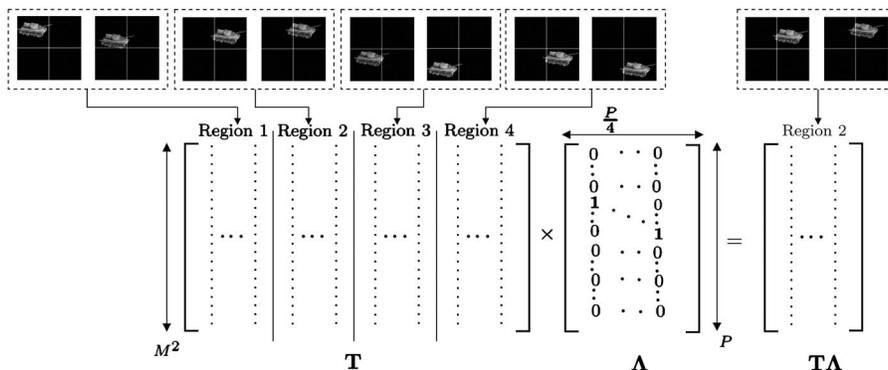
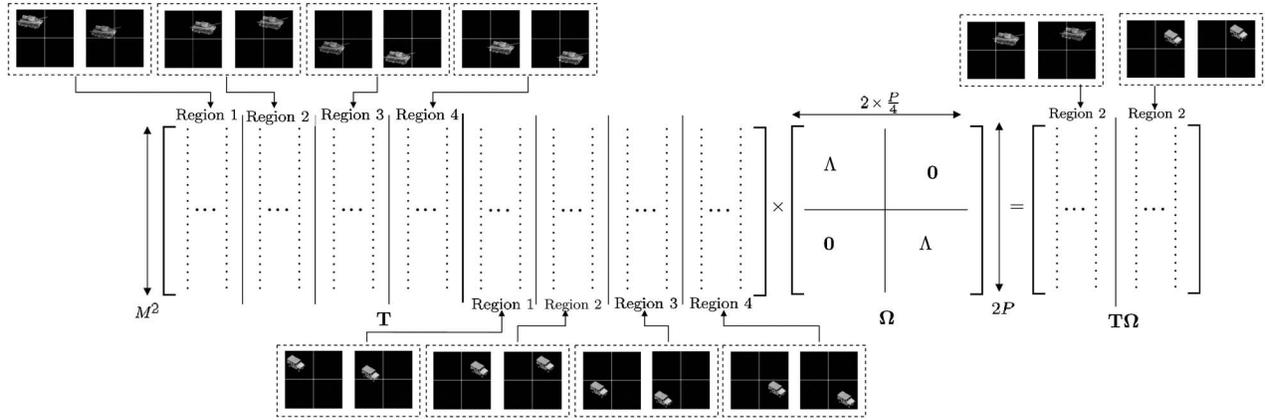


Fig. 8. Structure of \mathbf{T} and Λ matrices for the joint detection/localization problem.


 Fig. 9. Structure of \mathbf{T} and $\mathbf{\Omega}$ matrices for the joint classification/localization problem.

source X' and the term $\mathbf{T}\mathbf{\Lambda}(X)\vec{\rho}\alpha$, respectively. The channel operator \mathcal{H} is replaced with \mathbf{H} and \vec{N} is replaced by \vec{N}_c . Therefore, TSI and $mmse_H$ for this task can be expressed as

$$\text{TSI} = I(X'; \vec{R}) = \frac{1}{2} \int_0^s mmse_H(s') ds', \quad (24)$$

where

$$mmse_H(s) = \text{Tr}(\mathbf{H}^\dagger \Sigma_{\vec{N}_c}^{-1} \mathbf{H} (\mathbf{E}_{\vec{Y}} - \mathbf{E}_{\vec{Y}|X'})), \quad (25)$$

$$X' \in \{X, 0\}, \quad \vec{Y} = \mathbf{T}\mathbf{\Lambda}(X)\vec{\rho}\alpha. \quad (26)$$

The $(Q+1)$ -ary nature of the virtual source variable in the joint detection/localization task increases the upper bound on TSI as compared with that for the simple detection task. For the probabilities $\Pr(\alpha=1)=p$ and $\Pr(\alpha=0)=1-p$, the TSI is upper bounded by

$$J(X') = -(1-p)\log(1-p) - \sum_{q=1}^Q \Pr(X=q)\log \Pr(X=q), \quad (27)$$

where $\sum_{q=1}^Q \Pr(X=q)=p$. For the case of $p=\frac{1}{2}$ and $\Pr(X=q)=p/Q$, the maximum TSI is $[1+\frac{1}{2}\log Q]$ bits.

Finally, we consider the joint classification/localization task where the task of interest is to identify one of the two targets from H_1 or H_2 and localize it in one of Q regions. The exact position of the target within each region remains a nuisance parameter. The imaging model for this task is given by

$$\vec{R} = \sqrt{s}\mathbf{H}\mathbf{T}\mathbf{\Omega}(X)\boldsymbol{\rho}\vec{\alpha} + \vec{N}_c. \quad (28)$$

This model is the same as the one given in Eq. (23) except for minor modifications. The total number of positions that each target can take remains unchanged. However, now \mathbf{T} has dimensions $M^2 \times 2P$ and is given by $\mathbf{T} = [\mathbf{T}_{H_1} \quad \mathbf{T}_{H_2}]$, where \mathbf{T}_{H_i} is the target profile matrix for target i . The arrangement of the target profiles in \mathbf{T}_{H_1} and \mathbf{T}_{H_2} is similar to the arrangement described in Subsection 2.C. The virtual source in this case is $2Q$ -ary and given by $\vec{X}' = [X, \vec{\alpha}]$, where $X \in \{1, 2, \dots, Q\}$ indicates the region and $\vec{\alpha} \in \{[1, 0]^T, [0, 1]^T\}$ represents one of the two

targets. The localization matrix $\mathbf{\Omega}(X=i)$ now has dimensions $2P \times 2P_i$ for selecting the H_1 and H_2 profiles in the region i and is given by

$$\mathbf{\Omega}(X=i) = \begin{bmatrix} \mathbf{\Lambda}(X=i) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}(X=i) \end{bmatrix}, \quad (29)$$

where matrices $\mathbf{\Lambda}(X=i)$ and $\mathbf{0}$ are of dimension $P \times P_i$. The matrix $\mathbf{\Lambda}(X)$ is identical to the one in Eq. (22). Figure 9 illustrates the role of $\mathbf{T}\mathbf{\Omega}(X)$ in choosing the H_1 and H_2 profiles at all positions in the region specified by X . This example uses $X=2$, $Q=4$, and $P_i=P/4$ for $i=\{1, 2, 3, 4\}$. The matrix $\mathbf{T}\mathbf{\Omega}(X)$ in Eq. (28) is postmultiplied by the matrix $\boldsymbol{\rho}$ of dimension $2P_i \times 2$ to yield the targets H_1 and H_2 at one of the positions in region i . Here $\boldsymbol{\rho}$ is defined as

$$\boldsymbol{\rho} = \begin{bmatrix} \vec{\rho}_H & \mathbf{0} \\ \mathbf{0} & \vec{\rho}_H \end{bmatrix}, \quad (30)$$

where $\mathbf{0}$ is an all-zero P_i -dimensional column vector and $\vec{\rho}_H \in \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_{P_i}\}$, where \vec{e}_k is an indicator vector as before. Therefore, for $\vec{\rho}_H = \vec{e}_k$, $\mathbf{T}\mathbf{\Omega}(X)\boldsymbol{\rho}$ results in an $M^2 \times 2$ matrix with its first column representing H_1 at the k th position in region i and its second column representing H_2 at this same position. This result is then multiplied by $\vec{\alpha}$, which selects either H_1 or H_2 for $\vec{\alpha} = [1, 0]^T$ or $\vec{\alpha} = [0, 1]^T$, respectively.

The TSI expression in Eq. (24) requires only minor modifications to remain valid for the joint classification and localization problem. The upper bound for TSI in this task is given by

$$J(\vec{X}') = - \sum_{i=1}^2 \sum_{q=1}^P \Pr(X=q, \vec{\alpha}_i) \log \Pr(X=q, \vec{\alpha}_i), \quad (31)$$

where $\vec{\alpha}_1 = [0, 1]^T$, $\vec{\alpha}_2 = [1, 0]^T$, $\sum_{q=1}^Q \Pr(X=q, \vec{\alpha}_1) = 1-p$, and $\sum_{q=1}^Q \Pr(X=q, \vec{\alpha}_2) = p$. For the case when $p=\frac{1}{2}$, $\Pr(X=q, \vec{\alpha}_1) = (1-p)/Q$ and $\Pr(X=q, \vec{\alpha}_2) = p/Q$, the maximum TSI is $[1+\log Q]$ bits.

3. SIMPLE IMAGING EXAMPLES

The TSI framework described in the previous section allows us to evaluate the task-specific performance of an imaging system for a task defined by a specific encoding operator and virtual source variable. Three encoding op-

erators corresponding to three different tasks: (a) detection, (b) classification, and (c) joint detection/classification and localization have been defined. Now we apply the TSI framework to evaluate the performance of both a geometric imager and a diffraction-limited imager on these three tasks.

We begin by describing the source, object, and clutter used in the scene model. The source variable X in the detection task represents “tank present” or “tank absent” conditions with equal probability, i.e., $p = \frac{1}{2}$. In the classification task, the source variable \bar{X} represents “tank present” or “jeep present” states with equal probability. The joint localization task adds the position parameter to both the detection and classification tasks. From Eq. (16) we see that the source parameter is the input to the encoding operator, which in turn generates a scene consisting of both object and clutter. Here the scene \bar{Y} is of dimension 80×80 pixels ($M=80$). The object in the scene can be either a tank or a jeep at one of 64 equally likely positions ($P=64$). Therefore, the matrix \mathbf{T} has dimensions of 6400×64 for the detection task and 6400×128 for the classification task. In our scene model, the number of clutter components is set to $K=6$. Recall that clutter components are arranged as column vectors in the clutter matrix \mathbf{V}_c . Clutter is generated by combining these components, with relative weights specified by the column vector $\bar{\beta}$. Note that each clutter vector is nonrandom but the weight vector $\bar{\beta}$ follows a multivariate Gaussian distribution. In the simulation study the mean of $\bar{\beta}$ is set to $\bar{\mu}_{\bar{\beta}} = [160 \ 80 \ 40 \ 40 \ 64 \ 40]$ and covariance to $\bar{\Sigma}_{\bar{\beta}} = \bar{\mu}_{\bar{\beta}}^T \mathbf{I} / 5$. The clutter-to-noise ratio, denoted by c , is set to 1. The noise \bar{N} is zero mean with unity covariance matrix $\bar{\Sigma}_{\bar{N}} = \mathbf{I}$.

Monte Carlo simulations with importance sampling are used to estimate $mmse_H$ using the conditional mean estimators for a given task. The $mmse_H$ estimates are numerically integrated to obtain TSI over a range of s . For each value of s , we use 160,000 clutter and noise realizations in the Monte Carlo simulations.

A. Ideal Geometric Imager

The geometric imager represents an ideal imaging system with no blur, and therefore we set $\mathbf{H} = \mathbf{I}$. Figure 10 shows some example scenes resulting from object realizations measured in the presence of noise. Note that the object in the scene is either a tank or a jeep at one of the 64 positions.

We begin by describing the results for the detection task. Figure 11(a) and 11(b) show the plots of $mmse_H$ and TSI versus s , respectively. Recall that the $mmse_H$ is equal to the difference of $\mathbf{E}_{\bar{Y}}$ and $\mathbf{E}_{\bar{Y}|X}$, represented by the dotted and dashed curves, respectively, in Fig. 11(a). Also recall that $\mathbf{E}_{\bar{Y}|X}$ represents the $mmse$ in estimating \bar{Y} given the knowledge of both the measurement \bar{R} and source X . Therefore, we expect it to always be less than $\mathbf{E}_{\bar{Y}}$, which is the $mmse$ in estimating \bar{Y} given only the measurement \bar{R} . Figure 11(a) confirms this behavior. In the low- s region, $mmse_H$ (solid curve) is small, as both $\mathbf{E}_{\bar{Y}}$ and $\mathbf{E}_{\bar{Y}|X}$ are nearly equal. Despite the additional conditioning on X , $\mathbf{E}_{\bar{Y}|X}$ does not significantly improve upon $\mathbf{E}_{\bar{Y}}$, as the noise

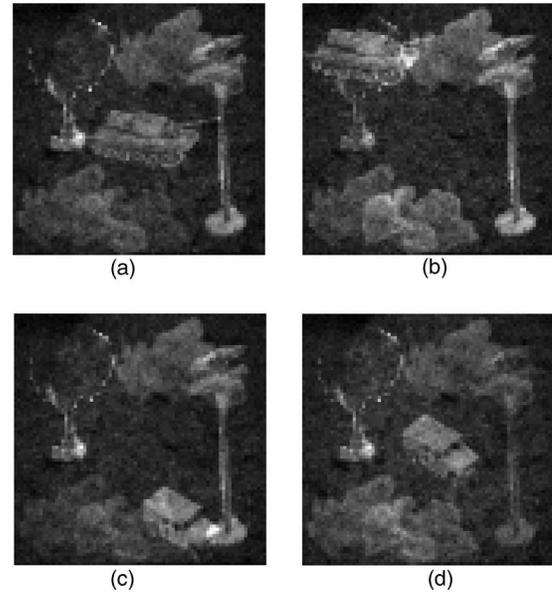


Fig. 10. Example scenes: (a) tank in the middle of the scene, (b) tank at the top of the scene, (c) jeep at the bottom of the scene, and (d) jeep in the middle of the scene.

remains the dominating factor. However, in the moderate- s region, $\mathbf{E}_{\bar{Y}|X}$ improves faster than $\mathbf{E}_{\bar{Y}}$ and therefore the $mmse_H$ increases here. In the high- s regime, the noise has a negligible effect and hence the additional knowledge of

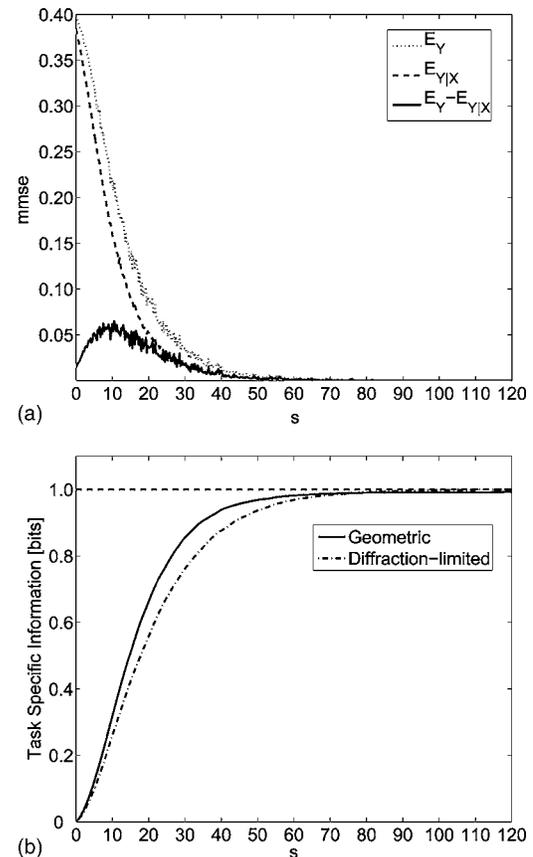


Fig. 11. Detection task: (a) $mmse$ versus signal-to-noise ratio for an ideal geometric imager and (b) TSI versus signal-to-noise ratio for geometric and diffraction-limited imagers.

X does not significantly improve $\mathbf{E}_{\tilde{Y}|X}$. This leads to the $mmse_H$ converging toward zero as both the $mmse$ components become equal. The solid curve in Fig. 11(b) shows the plot of TSI versus s . As expected, the TSI increases with s , eventually saturating at 1 bit. The saturation occurs because TSI is always upper bounded by the entropy of the virtual source X . The TSI plot confirms our expectations regarding blur-free imaging system performance with increasing s .

Now we consider TSI for the joint task of detecting and localizing a target. The scene is partitioned into four regions, i.e., $Q=4$. There are a total of 64 allowable target positions, with 16 positions in each region. Figure 12 shows some example scenes. Recall that the position of the target within each region is a nuisance parameter. We assume that the probability of a target being present or absent is $\frac{1}{2}$ and the conditional probability of a target in any of the four regions is $\frac{1}{4}$, given that the target is present. The entropy of the source variable therefore increases to 2 bits as per Eq. (27). Figure 13(a) shows a plot of $mmse$ versus s for the joint detection and localization task. The dotted curve represents the $mmse$ of the estimator conditioned over the image measurement only. The dashed curve corresponds to the $mmse$ of the estimator conditioned jointly on the virtual source variable and the image measurement. As expected, we see that $\mathbf{E}_{\tilde{Y}|X} \leq \mathbf{E}_{\tilde{Y}}$. The solid curve represents $mmse_H$, the difference between the dotted and dashed curves, and is integrated to yield TSI. The TSI of the geometric imager is plotted in a solid curve versus s in Fig. 13(b). We note that the TSI saturates at 2 bits as expected.

The previous two examples have demonstrated how the formalism of Section 2 can be applied to either a detection task or a joint detection/localization task. These examples have also confirmed two important TSI trends: (1) TSI is a monotonically increasing function of signal-to-noise ratio, and (2) TSI saturates at the entropy of the virtual source. Section 2 also described how a classification task or a joint classification/localization task may be captured

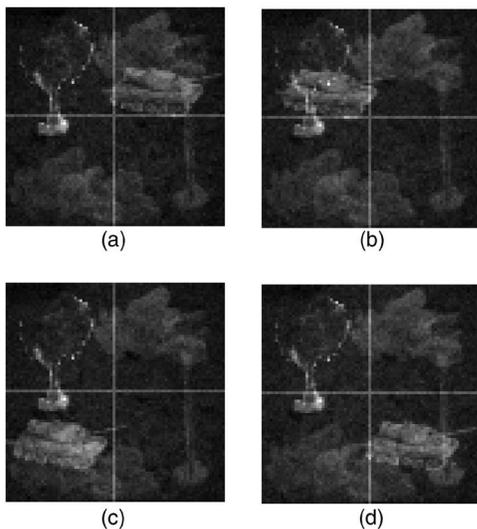


Fig. 12. Scene partitioned into four regions: (a) tank in the top-left region of the scene, (b) tank in the top-right region of the scene, (c) tank in the bottom-left region of the scene, and (d) tank in the bottom-right region of the scene.

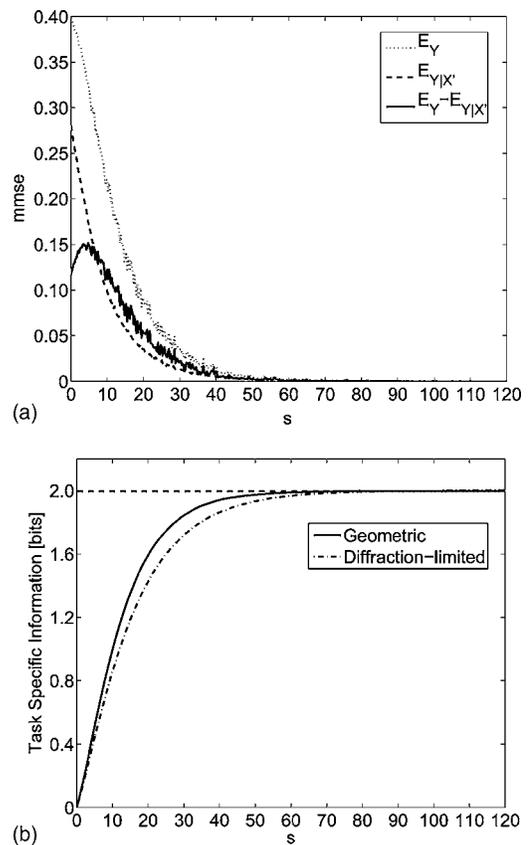


Fig. 13. Joint detection/localization task: (a) $mmse$ versus signal-to-noise ratio for an ideal geometric imager and (b) TSI versus signal-to-noise ratio for geometric and diffraction-limited imagers.

within the TSI formalism. The solid curve in Fig. 14 depicts the TSI obtained from an ideal geometric imager for a classification task in which the two classes are equally probable. Recall that for the classification task we treat the position as the nuisance parameter and so the equiprobable assumption results in a virtual source entropy of 1 bit. As expected, the TSI in Fig. 14 saturates at 1 bit. Figure 15 presents the results of the TSI analysis of the joint classification/localization task. Once again we have used two equally probable targets and $Q=4$ equally probable regions, resulting in a source entropy of 3 bits. We

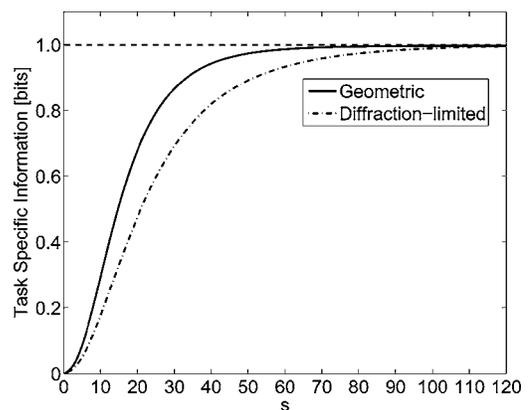


Fig. 14. Classification task: TSI versus signal-to-noise ratio for geometric and diffraction-limited imagers.

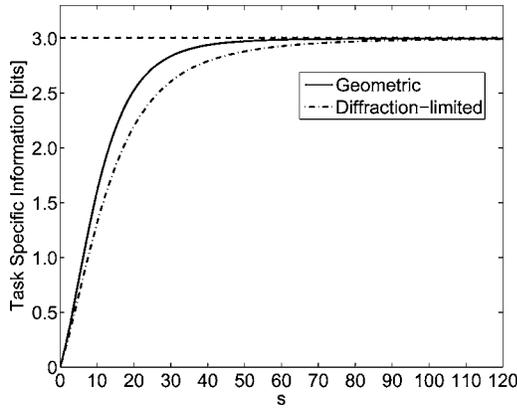


Fig. 15. Joint classification/localization task: TSI versus signal-to-noise ratio for geometric and diffraction-limited imagers.

see that once again, despite the measurement entropy that results from random clutter and noise, the TSI provides an accurate estimate of the task-relevant information, saturating at 3 bits.

B. Ideal Diffraction-Limited Imager

The previous subsection presented the TSI results for an ideal geometric imager. Those results should therefore be interpreted as upper bounds on the performance of any real-world imager. In this subsection we will examine the effect of optical blur on TSI. We will assume aberration-free, space-invariant, diffraction-limited performance. The discretized optical point-spread function (PSF) associated with a rectangular pupil can be expressed as [24]

$$h_{i,j} = \int_{-\Delta/2}^{\Delta/2} \int_{-\Delta/2}^{\Delta/2} \text{sinc}^2\left(\frac{x-i\Delta}{W}\right) \text{sinc}^2\left(\frac{y-j\Delta}{W}\right) dx dy, \quad (32)$$

where Δ is the detector pitch and W quantifies the degree of optical blur associated with the imager. Lexicographic ordering of this two-dimensional PSF yields one row of \mathbf{H} , and all other rows are obtained by lexicographically ordering shifted versions of this PSF. The optical blur is set to $W=2$, and the detector pitch is set to $\Delta=1$ so that the optical PSF is sampled at the Nyquist rate. The clutter and noise statistics remain unchanged.

Figure 16 shows examples of images that demonstrate the effects of both optical blur and noise. The object, as before, is either a tank or a jeep at one of the 64 positions. The plots of TSI versus s are represented by dashed-dotted curves for the detection and classification tasks in Figs. 11(b) and 14, respectively. The TSI metric verifies that imager performance is degraded due to optical blur compared with the geometric imager. For example, in the detection task, $s=34$ yields TSI=0.9 bit for the geometric imager, whereas a higher signal-to-noise ratio $s=43$ is required to achieve the same TSI for the diffraction-limited imager.

The dashed-dotted curves in Figs. 13(b) and 15 show the TSI versus s plots for the joint detection/localization and classification/localization tasks, respectively. Once again, we see that TSI is reduced due to optical blur. In Fig. 13(b) TSI=1.8 bit is achieved at $s=35$ for the diffraction-limited imager as opposed to $s=28$ in the case

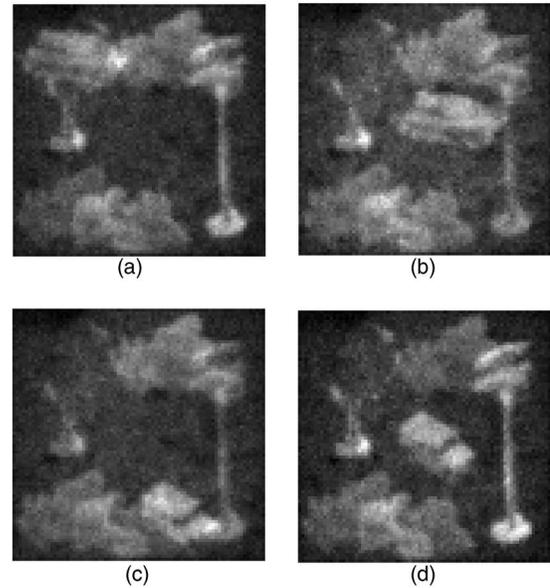


Fig. 16. Example scenes with optical blur: (a) tank at the top of the scene, (b) tank in the middle of the scene, (c) jeep at the bottom of the scene, and (d) jeep in the middle of the scene.

of the geometric imager for the detection/localization task. Similarly, for the classification/localization task the signal-to-noise ratio required to achieve TSI=2.7 bit increases by 10 due to the optical blur in the diffraction-limited imager.

In this section we have presented several numerical examples that demonstrate how the TSI analysis can be applied to various tasks and/or imaging systems. The results obtained herein are consistent with our expectations that (1) TSI increases with increasing signal-to-noise ratio, (2) TSI is upper bounded by $J(X)$, and (3) blur degrades TSI. Although these general trends were known in advance of our analysis, we are encouraged by our ability to quantify these trends using a formal approach. In the next section we will use a TSI analysis to evaluate the target-detection performance of two candidate projective imagers.

4. PROJECTIVE IMAGER

For task-specific applications (e.g., detection) an isomorphic measurement (i.e., a pretty picture) may not represent an optimal approach for extracting TSI in the presence of detector noise and a fixed photon budget. The dimensionality of the measurement vector has a direct effect on the measurement signal-to-noise ratio [25]. Therefore, we strive to design an imager that directly measures the scene information most relevant to the task while minimizing the number of detector measurements and thereby increasing the measurement signal-to-noise ratio. One approach toward this goal is to measure linear projections of the scene, yielding as many detector measurements as there are projections. We refer to such an imager as a projective imager. Figure 17 shows the imaging chain block diagram modified to include a projective transformation \mathcal{P} . For the projective imager the measurement can be written as

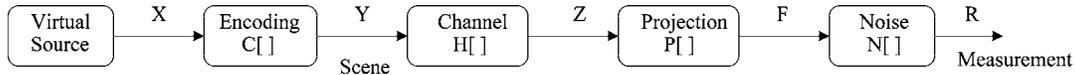


Fig. 17. Block diagram of a projective imager.

$$R = \mathcal{N}(\mathcal{P}\{\mathcal{H}[C(X)]\}). \quad (33)$$

We consider only discrete linear projections here; therefore the \mathcal{P} operator is represented by the matrix \mathbf{P} . If we consider the detection task from Subsection 2.B, then the measurement model for the projective imager can be written as

$$\vec{R} = \sqrt{s}\mathbf{PHT}\vec{\rho}X + \vec{N}'_c, \quad (34)$$

where

$$\vec{N}'_c = \sqrt{c}\mathbf{PHV}_c\beta + \vec{N}.$$

The TSI and the $mmse_H$ expressions for the projective imager are found by substituting \mathbf{PH} for \mathbf{H} in Eqs. (18)–(25), yielding

$$\text{TSI} \equiv I(X; \vec{R}) = \frac{1}{2} \int_0^s mmse_H(s') ds', \quad (35)$$

where

$$mmse_H(s) = \text{Tr}(\mathbf{H}^\dagger \mathbf{P}^\dagger \Sigma_{\vec{N}'_c}^{-1} \mathbf{PH}(\mathbf{E}_{\vec{Y}} - \mathbf{E}_{\vec{Y}|X})). \quad (36)$$

Here $\vec{Y} = \mathbf{T}\vec{\rho}X$ and $\mathbf{E}_{\vec{Y}}$ and $\mathbf{E}_{\vec{Y}|X}$ are given earlier in Eq. (10).

Similarly, for the joint detection/localization task from Subsection 2.D, the modified expressions for the imaging model and TSI are given by

$$\vec{R} = \sqrt{s}\mathbf{PHT}\Lambda(X)\vec{\rho}\alpha + \vec{N}'_c, \quad (37)$$

$$\text{TSI} \equiv I(X'; \vec{R}) = \frac{1}{2} \int_0^s mmse_H(s') ds', \quad (38)$$

where

$$mmse_H(s) = \text{Tr}(\mathbf{H}^\dagger \mathbf{P}^\dagger \Sigma_{\vec{N}'_c}^{-1} \mathbf{PH}(\mathbf{E}_{\vec{Y}} - \mathbf{E}_{\vec{Y}|X'})). \quad (39)$$

Here $X' \in \{X, 0\}$ and $\vec{Y} = \mathbf{T}\Lambda(X)\vec{\rho}\alpha$.

We consider projective imagers based on two classes of projection: (a) PC projections and (b) MF projections. Their performance is compared with that of the conventional diffraction-limited imager.

A. Principal Component Projection

PC projections are determined by the statistics of the object ensemble. For a set of objects O , the PC projections are defined as the eigenvectors of the object autocorrelation matrix R_{OO} given by

$$R_{OO} = \mathbb{E}(oo^T), \quad (40)$$

where $o \in O$ is a column vector formed by lexicographically arranging the elements of a two-dimensional object. Note that the expectation is over all objects in the set O . These PC projection vectors are used as rows of the projection matrix \mathbf{P}^* . In our numerical study, example objects

in the set O are obtained by generating sample realization of random scenes with varying clutter levels, target strength, and target position. Here we use 10,000 such object realizations to estimate R_{OO} . The projection matrix \mathbf{P}^* consists of F rows of length $M^2=6400$, which are the eigenvectors of R_{OO} corresponding to the F dominant eigenvalues. To ensure a fair comparison of the projective imager with the diffraction-limited imager, we constrain the total number of photons used by the former to be less than or equal to the total number photons used by the latter. The following normalization is applied to \mathbf{P}^* to enforce this photon constraint, resulting in the projection matrix \mathbf{P} ;

$$\mathbf{P} = \frac{1}{cs} \mathbf{P}^*, \quad (41)$$

where the maximum column sum $cs = \max_j \sum_{i=1}^F |\mathbf{P}^*|_{ij}$.

Figure 18 shows the TSI for this projective imager plotted as a function of s for the detection task. The dashed-dotted curve represents the TSI for the diffraction-limited imager from Subsection 3.B. Note that the TSI for a projective imager increases as the number of PC projections F is increased from 8 to 24. This can be attributed to the reduction in truncation error associated with increasing F . However, there is also an associated signal-to-noise ratio cost with increasing F , as we distribute the fixed photon budget across more measurements while the detector noise variance remains fixed. This effect is illustrated by the case $F=32$, where the TSI begins to deteriorate. This is especially evident at a low signal-to-noise ratio. Notwithstanding this effect, the PC projective imager is seen to provide improved task-specific performance compared with a conventional diffraction-limited imager, especially at a low signal-to-noise ratio. For example, the projective imager with $F=24$ achieves a TSI=0.9 bit at $s=18$, whereas the diffraction-limited imager requires $s=34$ to achieve the same TSI performance.

The TSI plot for the joint detection/localization task is shown in Fig. 19 for both the projective and diffraction-

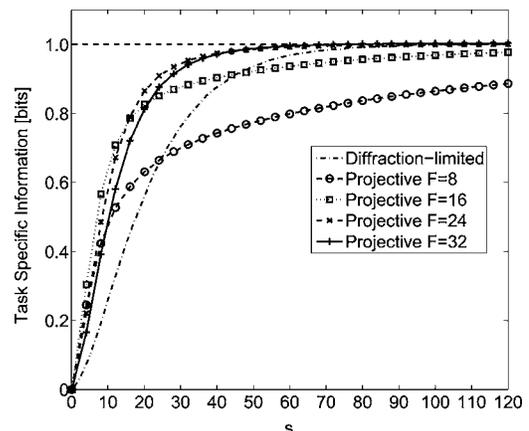


Fig. 18. Detection task: TSI for PC projective imager versus signal-to-noise ratio.

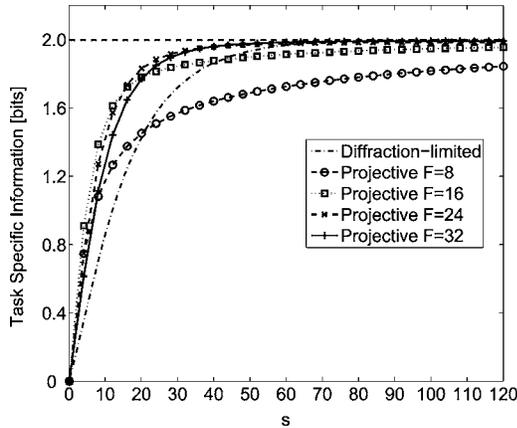


Fig. 19. Joint detection/localization task: TSI for PC projective imager versus signal-to-noise ratio.

limited imagers. We see the same trends as in Fig. 18. As before, a TSI rollover occurs at $F=32$ due to the signal-to-noise ratio trade-off associated with increasing F . In comparison with the diffraction-limited imager that requires $s=35$ to achieve TSI=1.8 bit, the projective imager with $F=24$ achieves the same level of performance at $s=19$.

Although we have shown that the PC projective imager provides larger TSI than the diffraction-limited imager, we cannot claim that the PC projections are an optimal choice. This is because PC projections seek to minimize the reconstruction error toward the goal of estimating the whole scene [15], which is an overly stringent requirement for a detection task. In fact, for a detection problem it is well known that the generalized MF approach is optimal in terms of the Neyman–Pearson criterion [26]. In the next section we present the TSI results for a MF projective imager.

B. Matched-Filter Projection

For a detection problem in which both the signal and background are known, the generalized MF provides the optimal performance in terms of maximizing the probability of detection for a fixed false alarm rate [26]. Recall that in our detection problem the target position is a nuisance parameter that must be estimated implicitly. In such a case, instead of a MF (e.g., correlator), we consider a set of matched projections. Each matched projection corresponds to the target at a given position. Therefore the resulting projective imager yields the inner product between the scene and the target at a particular position specified by each projection. Note that projective imaging in such a case is similar to an optical correlator except that in an optical correlator the inner-product values are obtained for all possible shifts of the target: Our projective imager will compute inner products for only a subset of these shifts.

The projection matrix \mathbf{P} of the matched projection imager is defined as

$$\mathbf{P} = \bar{\mathbf{T}} \Sigma_{N_c}^{-1}, \quad (42)$$

where $\bar{\mathbf{T}}$ is the modified target profile matrix with each row corresponding to a target profile at a specific position. The number of positions chosen is F , and therefore the di-

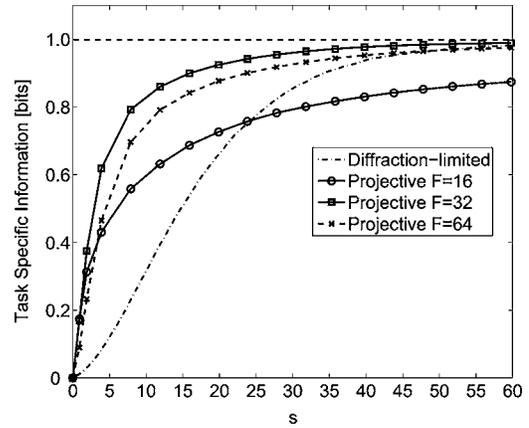


Fig. 20. Detection task: TSI for MF projective imager versus signal-to-noise ratio.

mensions of the matrix $\bar{\mathbf{T}}$ is $F \times M^2$. The target positions for constructing $\bar{\mathbf{T}}$ are chosen such that they are equally spaced with some overlap between the profiles at the adjacent positions. The target profile matrix $\bar{\mathbf{T}}$ is postmultiplied by $\Sigma_{N_c}^{-1}$, resulting in the projection matrix \mathbf{P} , in order to account for the effects of detector noise [26]. The dimensions of \mathbf{P} are $F \times M^2$. Therefore, the projective imager with projection \mathbf{P} yields F measurements as opposed to M^2 measurements as in the case of the diffraction-limited imager, where $F \ll M^2$. As in the previous section, the MF projection matrix \mathbf{P} is normalized as per Eq. (41) to allow for a fair comparison with the diffraction-limited imager.

Recall that the target can appear at one of the 64 possible positions; hence $F=64$ is the maximum number of projections. Figure 20 shows the plot of TSI versus s for the MF projective imager with $F=16$, 32, and 64. As before, we see that TSI increases with the number of projections F . However, at $F=32$ the TSI shows the rollover effect due to the signal-to-noise-ratio cost associated with increasing F . Ideally, we expect that the maximum TSI is obtained for $F=64$, as it includes all possible target positions. However, there is some overlap between the target profiles at adjacent positions and so $F \leq 64$ projections are sufficient to extract the detection-task-related information. Note that, in the absence of the photon-count con-

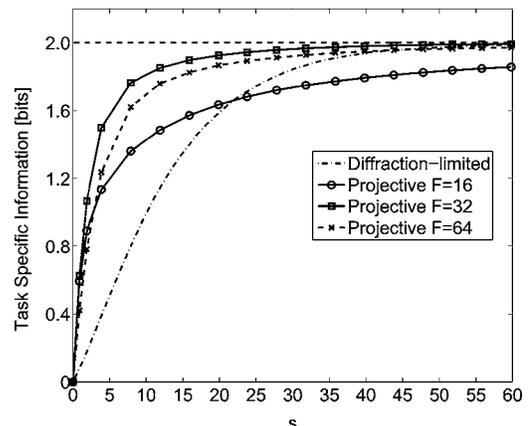


Fig. 21. Joint detection/localization task: TSI for MF projective imager versus signal-to-noise ratio.

straint, a choice of $F=64$ would have indeed provided the highest TSI. As expected, the MF projection imager yields a better performance compared with the PC projection imager. For example, to achieve $\text{TSI}=0.9$ bit, the MF projections with $F=32$ requires $s=17$ compared with $s=23$ for the PC projections.

The TSI versus s plot for the joint detection/localization task is shown in Fig. 21. Similar to the detection task, we observe the rollover effect at $F=32$. As expected, TSI saturates at 2 bits at high signal-to-noise ratio. The MF projective imager with $F=32$ offers improved performance, achieving $\text{TSI}=1.8$ bit at $s=8$ compared with $s=19$ required by the PC projection imager.

5. CONCLUSIONS

Imagery is often employed to accomplish some computational task. In such cases the task-specific performance becomes crucial. The TSI content of an image measurement can serve as an objective measure of the imaging system performance in such a case. In this paper, we have proposed a framework for the definition of TSI in terms of the well-known Shannon mutual information measure. The use of the virtual source variable is key to our definition of TSI and, to our knowledge, a unique method of embedding task specificity in the scene model itself. The recently discovered relationship between mutual information and *mmse* allows us to calculate the TSI from the simulated performance of conditional mean estimators. The proposed TSI framework is applied to evaluate the performance of geometric and diffraction-limited imaging systems for three tasks: detection, classification, and joint detection/classification and localization. The results obtained from the simulation study confirm our intuition about the performance of these two candidate imaging systems, thereby establishing TSI as an objective task-specific performance metric.

We also exercised the TSI framework to study the design of two projective/compressive imagers. In the case of the PC projective imager, we found that the TSI analysis confirmed the previously known trade-off with increasing number of projections. The TSI performance of the MF projective imager verified that it can be a superior projective imager design for a detection task. From these results we conclude that TSI is a useful metric for studying the task-specific performance of an imaging system. We note that TSI may serve as an upper bound on the performance of any algorithm that attempts to extract TSI from the measurement data.

The TSI framework presented here can be easily extended to continuous scenes with minor modifications to the estimators derived in Appendix A. Note that we have considered only an additive Gaussian noise model here; however, the generalized relation between mutual information and estimation theory recently reported in [27] will allow us to extend our model to include nonadditive and/or non-Gaussian cases in the future. We would also like to note that our nonideal additive clutter/target model can be modified to use a more realistic replacement model as defined in [28]. Currently, we are pursuing methods to reduce the computation complexity of estimating *mmse*. A reduction in computational complexity would

allow us to investigate more complex scene models. We note that TSI can be used as a metric to optimize an imaging system for maximizing the task-specific performance/information: an area of ongoing research in our group.

APPENDIX A: CONDITIONAL MEAN ESTIMATORS FOR DETECTION, CLASSIFICATION, AND LOCALIZATION TASKS

Here we derive explicit expressions for the conditional mean estimators $\mathbb{E}(\vec{Y}|\vec{R})$ and $\mathbb{E}(\vec{Y}|\vec{R},\vec{X})$ for each of the three tasks: detection, classification, and localization. We define $\mathbb{E}(\vec{Y}|\vec{R})$ as the expected value of \vec{Y} given the measurement \vec{R} and write it as

$$\begin{aligned} \mathbb{E}(\vec{Y}|\vec{R}) &= \sum_l \vec{Y}_l \Pr(\vec{Y} = \vec{Y}_l|\vec{R}) \\ &= \sum_l \vec{Y}_l \frac{pr(\vec{R}|\vec{Y} = \vec{Y}_l)\Pr(\vec{Y} = \vec{Y}_l)}{\sum_m pr(\vec{R}|\vec{Y} = \vec{Y}_m)\Pr(\vec{Y} = \vec{Y}_m)}, \end{aligned} \quad (\text{A1})$$

where \vec{Y}_l spans over all the possible scenes that can be generated by the random encoding function. Recall that for the detection task defined in Subsection 2.B, $\vec{Y} = \mathbf{T}\vec{\rho}X$ and virtual source X is binary. Therefore, for $X=1$, \vec{Y} can take P different values corresponding to the P possible positions; \vec{Y} is equal to zero for $X=0$. We define $\Pr(X=1) = p$, $\Pr(X=0) = 1-p$, $\Pr(\vec{Y}=\vec{0}) = 1-p$, and $\Pr(\vec{Y}=\vec{Y}_l) = p/P$, where $l = \{1, 2, \dots, P\}$. Substituting these probabilities into (A1), we obtain

$$\mathbb{E}(\vec{Y}|\vec{R}) = \frac{\sum_{l=1}^P p \vec{Y}_l pr(\vec{R}|\vec{Y} = \vec{Y}_l)}{\sum_{m=1}^P p [pr(\vec{R}|\vec{Y} = \vec{Y}_m)] + (1-p)P[pr(\vec{R}|\vec{Y} = \vec{0})]}. \quad (\text{A2})$$

Here the conditional probability density function $pr(\vec{R}|\vec{Y} = \vec{Y}_l)$ is Gaussian and is given by

$$\begin{aligned} pr(\vec{R}|\vec{Y} = \vec{Y}_l) &= \frac{1}{(2\pi)^{K/2} \sqrt{\det \Sigma_{\vec{R}|\vec{Y}}} \\ &\times \exp \left[-\frac{1}{2} (\Theta_1 + \Theta_{2l} + \Theta_{3l} + \Theta_4) \right], \end{aligned} \quad (\text{A3})$$

where

$$\Sigma_{\vec{R}|\vec{Y}} = c \cdot \mathbf{PHV}_c \Sigma_{\vec{\beta}} (\mathbf{PHV}_c)^T + \sigma_N^2 \mathbf{I}, \quad (\text{A4})$$

$$\Theta_1 = \vec{R}^T \Sigma_{\vec{R}|\vec{Y}}^{-1} \vec{R} - 2\sqrt{c} \cdot \vec{R}^T \Sigma_{\vec{R}|\vec{Y}}^{-1} \mathbf{PHV}_c \vec{\mu}_{\vec{\beta}}, \quad (\text{A5})$$

$$\Theta_{2l} = -2\sqrt{s} \cdot \vec{R}^T \Sigma_{\vec{R}|\vec{Y}}^{-1} \mathbf{P} \mathbf{H} \vec{Y}_l, \quad (\text{A6})$$

$$\begin{aligned} \Theta_{3l} = & s \cdot \vec{Y}_l^T \mathbf{H}^T \mathbf{P}^T \Sigma_{\vec{R}|\vec{Y}}^{-1} \mathbf{P} \mathbf{H} \vec{Y}_l \\ & + 2\sqrt{s \cdot c} \cdot \vec{Y}_l^T \mathbf{H}^T \mathbf{P}^T \Sigma_{\vec{R}|\vec{Y}}^{-1} \mathbf{P} \mathbf{H} \mathbf{V}_c \vec{\mu}_{\vec{\beta}}, \end{aligned} \quad (\text{A7})$$

$$\Theta_4 = c \cdot \vec{\mu}_{\vec{\beta}}^T \mathbf{V}_c^T \mathbf{H}^T \mathbf{P}^T \Sigma_{\vec{R}|\vec{Y}}^{-1} \mathbf{P} \mathbf{H} \mathbf{V}_c \vec{\mu}_{\vec{\beta}}. \quad (\text{A8})$$

Substituting Eq. (A3) into (A2) and simplifying yields the following expression:

$$\mathbb{E}(\vec{Y}|\vec{R}) = \frac{\sum_{l=1, \vec{X}=[1,0]^T}^P p \vec{Y}_l \cdot \exp\left[-\frac{1}{2}(\Theta_{2l} + \Theta_{3l})\right] + \sum_{l=1, \vec{X}=[0,1]^T}^P (1-p) \vec{Y}_l \cdot \exp\left[-\frac{1}{2}(\Theta_{2l} + \Theta_{3l})\right]}{\sum_{m=1, \vec{X}=[1,0]^T}^P p \exp\left[-\frac{1}{2}(\Theta_{2m} + \Theta_{3m})\right] + \sum_{m=1, \vec{X}=[0,1]^T}^P (1-p) \exp\left[-\frac{1}{2}(\Theta_{2m} + \Theta_{3m})\right]}. \quad (\text{A10})$$

This expression is similar to Eq. (A9) except for about twice as many terms in the numerator and the denominator.

For the joint detection and localization task, the estimator $\mathbb{E}(\vec{Y}|\vec{R})$ can be obtained by minor modifications to Eq. (A9). Considering the probabilities specified in Subsection 2.D, the modified expression can be found as

$$\mathbb{E}(\vec{Y}|\vec{R}) = \frac{\sum_{i=1}^Q \sum_{l=1}^{P_i} \frac{\Pr(X=i)}{P_i} \vec{Y}_{i,l} \cdot \exp\left[-\frac{1}{2}(\Theta_{i,2l} + \Theta_{i,3l})\right]}{\sum_{j=1}^Q \sum_{m=1}^{P_j} \frac{\Pr(X=j)}{P_j} \cdot \exp\left[-\frac{1}{2}(\Theta_{j,2m} + \Theta_{j,3m})\right] + (1-p)}, \quad (\text{A11})$$

where $Y_{i,l}$ is the target profile at the l th position of region i . Here $\Theta_{i,2l}$ and $\Theta_{i,3l}$ are evaluated using Eqs. (A6) and (A7), respectively, by substituting Y_l with the corresponding $Y_{i,l}$. Similarly for the joint classification and localization task, the estimator $\mathbb{E}(\vec{Y}|\vec{R})$ can be written as

$$\mathbb{E}(\vec{Y}|\vec{R}) = \frac{\sum_{i=1}^Q \sum_{l=1}^{P_i} \sum_{\vec{\alpha}=[1,0]^T, [0,1]^T} \frac{\Pr(X=i, \vec{\alpha})}{P_i} \vec{Y}_{i,l, \vec{\alpha}} \cdot \exp\left[-\frac{1}{2}(\Theta_{i,2l} + \Theta_{i,3l})\right]}{\sum_{j=1}^Q \sum_{m=1}^{P_j} \sum_{\vec{\alpha}=[1,0]^T, [0,1]^T} \frac{\Pr(X=j, \vec{\alpha})}{P_j} \cdot \exp\left[-\frac{1}{2}(\Theta_{j,2m} + \Theta_{j,3m})\right]}, \quad (\text{A12})$$

where $Y_{i,l, \vec{\alpha}}$ is the target profile specified by $\vec{\alpha}$ at the l th position of region i . Here $\Theta_{i,2l}$ and $\Theta_{i,3l}$ are evaluated using Eqs. (A6) and (A7), respectively, by substituting Y_l by its respective $Y_{i,l, \vec{\alpha}}$.

Now we derive the expressions for the estimator $\mathbb{E}(\vec{Y}|\vec{R}, \vec{X})$ required in evaluating Eq. (18) for each task. The estimator is defined as

$$\mathbb{E}(\vec{Y}|\vec{R}, \vec{X}) = \sum_l \vec{Y}_l \cdot \Pr(\vec{Y} = \vec{Y}_l | \vec{R}, \vec{X}). \quad (\text{A13})$$

We may express the conditional probability $\Pr(\vec{Y} = \vec{Y}_l | \vec{R}, \vec{X})$ using Bayes's law as follows:

$$\mathbb{E}(\vec{Y}|\vec{R}) = \frac{\sum_{l=1}^P p \vec{Y}_l \cdot \exp\left[-\frac{1}{2}(\Theta_{2l} + \Theta_{3l})\right]}{\sum_{m=1}^P p \exp\left[-\frac{1}{2}(\Theta_{2m} + \Theta_{3m})\right] + (1-p)P}. \quad (\text{A9})$$

Note that here \mathbf{P} represents the projection matrix of the projective imager and must be assumed as \mathbf{I} otherwise.

Next consider the classification task specified in Subsection 2.C, where $\vec{Y} = \mathbf{T} \vec{\rho} \vec{X}$ and virtual source \vec{X} is binary. For \vec{X} equal to $[1, 0]^T$ or $[0, 1]^T$, \vec{Y} will have P possible values corresponding to as many positions. Assuming $\Pr(\vec{X} = [1, 0]^T) = p$, $\Pr(\vec{X} = [0, 1]^T) = 1-p$ and equiprobable positions for both targets, we obtain

$$\Pr(\vec{Y} = \vec{Y}_l | \vec{R}, \vec{X}) = \frac{pr(\vec{R}, \vec{X} | \vec{Y}_l) \Pr(\vec{Y} = \vec{Y}_l)}{pr(\vec{R}, \vec{X})} \quad (\text{A14})$$

$$= \frac{pr(\vec{R} | \vec{Y} = \vec{Y}_l, \vec{X}) \Pr(\vec{X} | \vec{Y} = \vec{Y}_l) \Pr(\vec{Y} = \vec{Y}_l)}{\sum_m pr(\vec{R} | \vec{Y} = \vec{Y}_m, \vec{X}) \Pr(\vec{X} | \vec{Y} = \vec{Y}_m) \Pr(\vec{Y} = \vec{Y}_m)}. \quad (\text{A15})$$

For the detection task in Subsection 2.B, the virtual source variable X is binary; therefore, by substituting Eqs. (A15) and (A3) into (A13) and simplifying, we obtain the following expressions for the estimator:

$$\mathbb{E}(\vec{Y}|\vec{R}, X=1) = \frac{\sum_{l=1}^P \vec{Y}_l \cdot \exp\left[-\frac{1}{2}(\Theta_{2l} + \Theta_{3l})\right]}{\sum_{m=1}^P \exp\left[-\frac{1}{2}(\Theta_{2m} + \Theta_{3m})\right]},$$

$$\mathbb{E}(\vec{Y}|\vec{R}, X=0) = \mathbf{0}, \quad (\text{A16})$$

where \vec{Y}_l in Eq. (A16) is the target profile at the l th position; Θ_{2l} and Θ_{3l} in Eq. (A16) are evaluated using Eqs. (A6) and (A7), respectively. Similarly, for the classification task defined in Subsection 2.C, the estimator in Eq. (A13) can be written as

$$\mathbb{E}(\vec{Y}|\vec{R}, \vec{X}) = \frac{\sum_{l=1}^P \vec{Y}_l \cdot \exp\left[-\frac{1}{2}(\Theta_{2l} + \Theta_{3l})\right]}{\sum_{m=1}^P \exp\left[-\frac{1}{2}(\Theta_{2m} + \Theta_{3m})\right]}, \quad (\text{A17})$$

where \vec{Y}_l in this case is the target profile specified by \vec{X} at the l th position.

Recall from Subsection 2.D that for the joint detection and localization task, the virtual source variable X' is $(Q+1)$ -ary. Note that $X'=X$, where X denotes the region in which the target is present when $\alpha=1$ and $X'=0$ when $\alpha=0$. The estimator in the Eq. (A13) for this case is given by

$$\mathbb{E}(\vec{Y}|\vec{R}, X=i, \alpha=1) = \frac{\sum_{l=1}^{P_i} \vec{Y}_{i,l} \cdot \exp\left[-\frac{1}{2}(\Theta_{i,2l} + \Theta_{i,3l})\right]}{\sum_{m=1}^{P_i} \exp\left[-\frac{1}{2}(\Theta_{i,2m} + \Theta_{i,3m})\right]},$$

$$\mathbb{E}(\vec{Y}|\vec{R}, \alpha=0) = \mathbf{0}, \quad (\text{A18})$$

where $X=i$ implies that the target is present in region i and $\vec{Y}_{i,l}$ is the target profile at the l th position of region i . Once again, $\Theta_{i,2l}$ and $\Theta_{i,3l}$ are evaluated using Eqs. (A6) and (A7), respectively, by substituting Y_l by the appropriate $Y_{i,l}$. In a similar manner the estimator $\mathbb{E}(\vec{Y}|\vec{R}, \vec{X})$ for the joint classification and localization task can be expressed as

$$\mathbb{E}(\vec{Y}|\vec{R}, X=i, \vec{\alpha}) = \frac{\sum_{l=1}^{P_i} \vec{Y}_{i,l,\vec{\alpha}} \cdot \exp\left[-\frac{1}{2}(\Theta_{i,2l} + \Theta_{i,3l})\right]}{\sum_{m=1}^{P_i} \exp\left[-\frac{1}{2}(\Theta_{i,2m} + \Theta_{i,3m})\right]}, \quad (\text{A19})$$

where $Y_{i,l,\vec{\alpha}}$, $\Theta_{i,2l}$, and $\Theta_{i,3l}$ have the same meaning as in Eq. (A12).

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support of the Defense Advanced Research Projects Agency's (DARPA's)

MONTAGE program under which this work was carried out. We would also like to thank the reviewers for suggesting several useful comments for making the paper more accessible.

REFERENCES

1. J. A. O'Sullivan, R. E. Blahut, and D. L. Snyder, "Information-theoretic image formation," *IEEE Trans. Image Process.* **44**, 2094–2123 (1998).
2. A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.* **15**, 23–50 (1998).
3. A. Ashok and M. A. Neifeld, "Information-based analysis of simple incoherent imaging systems," *Opt. Express* **11**, 2153–2162 (2003).
4. F. O. Huck, C. L. Fales, and Z. Rahman, "An information theory of visual communication," *Philos. Trans. R. Soc. London, Ser. A* **354**, 2193–2248 (1996).
5. F. O. Huck and C. L. Fales, "Information-theoretic assessment of sampled imaging systems," *Opt. Eng. (Bellingham)* **38**, 742–762 (1999).
6. J. Ahlberg and I. Renhorn, "An information-theoretic approach to band selection," *Proc. SPIE* **5811**, 15–23 (2005).
7. S. P. Awate, T. Tasdizen, N. Foster, and R. T. Whitaker, "Adaptive, nonparametric Markov modeling for unsupervised, MRI brain-tissue classification," *Med. Image Anal.* **10**, 726–739 (2006).
8. J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Process.* **10**, 1647–1658 (2001).
9. L. Zhen and Karam, "Mutual information-based analysis of JPEG2000 contexts," *IEEE Trans. Image Process.* **14**, 411–422 (2005).
10. D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice* (Springer, 2002).
11. T. Cover and J. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
12. H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," *J. Opt. Soc. Am. A* **7**, 1266–1278 (1990).
13. H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance," *J. Opt. Soc. Am. A* **12**, 834–852 (1995).
14. H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions," *J. Opt. Soc. Am. A* **15**, 1520–1535 (1998).
15. H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley-Interscience, 2004).
16. M. Tanner, *Tools for Statistical Inference*, 2nd ed. (Springer, 1993).
17. J. S. Liu, *Monte Carlo Strategies in Scientific Computing* (Springer, 2001).
18. C. P. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer, 2004).
19. A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice* (Springer, 2001).
20. L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory* **20**, 284–287 (1974).
21. D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory* **51**, 1261–1282 (2005).
22. H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory* (Wiley, 1968).
23. D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory* **52**, 141–154 (2006).

24. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, 1996), Chap. 7.
25. M. A. Neifeld and P. Shankar, "Feature-specific imaging," *Appl. Opt.* **42**, 3379–3389 (2003).
26. S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice Hall, 1993).
27. D. P. Palomar and S. Verdu, "Representation of mutual information via input estimates," *IEEE Trans. Inf. Theory* **53**, 453–470 (2007).
28. N. Towghi and B. Javidi, "Optimum receivers for pattern recognition in the presence of Gaussian noise with unknown statistics," *J. Opt. Soc. Am. A* **18**, 1844–1852 (2001).