Information Theoretic Characterizations of Coded Imaging-based Space Object Identification

Douglas A. Hope and Sudhakar Prasad
Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131
email: douglas.a.hope@gmail.com

Abstract

Obtaining information about the surface material composition of man-made objects such as earth orbiting satellites is an important goal of space object identification (SOI). Often initial information about the possible material composition of such targets is known *a priori*, at least in a statistical sense, in the form of a material database which includes the measured reflectance curves of materials most often used in construction of satellites. In fact, such a database serves as an appropriate sparse basis for modeling the target using a coded spectral imaging system. A coded imaging system exploits the fact that images of man-made objects posses correlations in their physical structure. These correlations, which can occur spatially as well as spectrally, can suggest a more natural sparse basis for compressing and representing the scene when compared to standard pixels or voxels. A coded imaging system attempts to acquire and encode the scene in this sparse-basis, while preserving all relevant information in the scene. Due to strong *a priori* statistical information in the form of the material database and knowledge of their standard morphology, statistical information provides a natural theoretical framework for assessing the content, acquisition, and processing of information by a coded imaging system about man-made space objects. Here, we will demonstrate the use of a statistical model of a coded imaging system for which we compute statistical information in the coded measurements using a highly efficient Monte-Carlo based algorithm [1]. Specifically, we will study the problem of encoding spatial information about the scene at different wavelengths into the measurements. An important parameter that will concern us is the minimum number of spectral measurements required to unambiguously identify each material in the scene, given some level of noise in the data. Due to our information-based approach, we have the ability to study the effects of gradually introducing more prior information via the material database on the identification of the materials from the measurements. We expect interdependence between the choice of spatial measurement vectors and the number of spectral channels. We anticipate that our information-theoretic results will yield insight into how to optimally choose the measurement vectors. Such knowledge will be crucial when designing a system to identify materials with similar reflectance curves.

1 Introduction

An important goal of SOI is the identification of man-made satellites. Successful identification depends on knowledge of the surface material components and their respective boundaries. One source of this information is ground-based images of the target. However, do to image blur caused by both the telescope and the turbulent atmosphere, these measurements only convey partial information about the materials and boundaries.

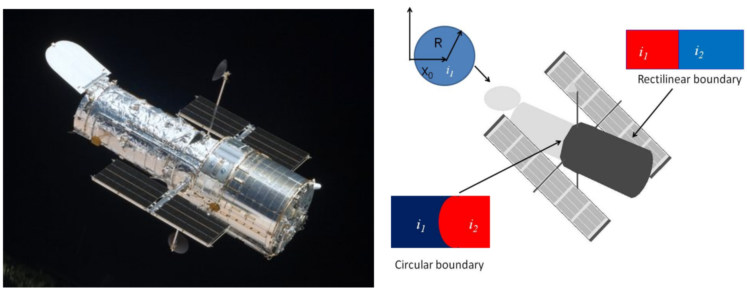
The ability of an electro-optical (EO) sensing system to measure and extract the most information possible from a limited set of measurements of the target is a crucial design requirement for future space surveillance systems. First, one must quantify the information content of a target and one approach is to introduce a model of the target in terms of shape primitives as shown in Fig. 1. The target can then be expressed as a vector of parameters, $\mathbf{X}$, for each shape primitive and its corresponding material with a probability distribution function (PDF) $p(\mathbf{x})$. The differential Shannon entropy [1], which is defined as the following self-average computed over this PDF,

$$h(\mathbf{X}) = -\int p(\mathbf{X}) \ln p(\mathbf{X}) d\mathbf{x} = -\langle \ln p_X \rangle, \tag{1}$$

is a measure of the information-carrying potential of the random variable X and represents the maximum information that can be encoded into the variable X. The more "states" that are accessible to the random variable X the larger its entropy and information carrying potential.

When modeled as a communication channel, a space surveillance system processes this input signal yielding a channel output $\mathbf{Y}$. The amount of information in $\mathbf{Y}$, that is conveyed about channel input $\mathbf{X}$ is referred to as the mutual information (MI) between X and Y and is defined as,

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &= H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}) \\ &= H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \\ &= H(\mathbf{X}) + H(\mathbf{Y}) - H(\mathbf{X},\mathbf{Y}), \end{aligned} \tag{2}$$



(a) (b)

Figure 1: **(Left) Hubble Space Telescope (Right) Model of HST using shape primitives with examples of local boundaries that accurately model the material boundaries on the target**

where the conditional entropies, $H(\mathbf{X}|\mathbf{Y})$ and $H(\mathbf{Y}|\mathbf{X})$, and the joint entropy $H(\mathbf{X},\mathbf{Y})$ are defined formally via relation (1) with the probability distribution P_X replaced by the corresponding conditional or joint PDFs inside the logarithm. The equivalence of the three forms in Eq.(2) is easily established by means of the Bayes theorem.

2 Local boundaries

The information measures in Eq. (1) describe the global information and, in most cases, are computational prohibitive. Alternately, one can compute information at local places in the target scene, which can in many cases be possible especially where the boundaries can be expressed in terms of a small set of parameters. This idealization is likely to be accurate locally in the vicinity of a boundary between two materials on a more realistic object surface, while affording a relatively simple information theoretic description of system performance as shown in Fig 2.

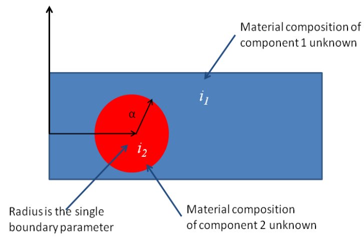


Figure 2: **Model of a local boundary that consist of two components, a circular component on a background surface. The circular component and the background each represent a surface material. The boundary of the circular component is described by the radius of the component.**

This local material boundary problem at wavelength λ has a spatial distribution,

$$f^{(j,\ell,\alpha)}(\mathbf{x},\lambda) = i^{(1,j)}(\lambda)q^{(1,\alpha)}(\mathbf{x}) + i^{(2,\ell)}(\lambda)q^{(2,\alpha)}(\mathbf{x}) \tag{3}$$

where $q_w^{(1,\alpha)}$ and $q_w^{(2,\alpha)}$ are the profile functions for the two material components that have a common boundary defined by α, the radius of the circular component. The quantities $i^{(1,j)}(\lambda)$ and $i^{(2,\ell)}(\lambda)$ represent the brightness at the individual pixels of the two materials (indexed by j and ℓ) at wavelength λ. The blurred image is a convolution of

$f^{(j,\ell,\alpha)}(\mathbf{x},\lambda)$ and the point-spread-function, $h_\lambda(\mathbf{x})$, of the telescope,

$$\begin{aligned} g^{(j,\ell,\alpha)}(\mathbf{x},\lambda) &= \sum_{x'} f^{(j,\ell,\alpha)}(\mathbf{x},\lambda) h_\lambda(\mathbf{x}'-\mathbf{x}) \\ &= i^{(1,j)}(\lambda)\left(q^{(1,\alpha)}\odot h_\lambda\right)(\mathbf{x}) + i^{(2,\ell)}(\lambda)\left(q^{(2,\alpha)}\odot h_\lambda\right)(\mathbf{x}). \end{aligned} \tag{4}$$

We compute the observed image in a spectral band (indexed by subscript w) of width $\Delta\lambda$ centered at wavelength λ_0 by integrating over the wavelengths in the band,

$$g_w^{(j,\ell,\alpha)}(\mathbf{x}) = \int_{\lambda_0-\Delta\lambda/2}^{\lambda_0+\Delta\lambda/2} W(\lambda) g^{(j,\ell,\alpha)}(\mathbf{x},\lambda) d\lambda, \tag{5}$$

where $W(\lambda)$ denotes the transmittance function of the filter. In this paper we will take $W(\lambda)$ to be constant over the passband of the filter.

The typical compressive sensing (CS) approach [7] requires making projective measurements using K spatial measurement vectors (masks) defined in the focal plane of the imaging system. The type of mask is known in advance of the measurement and is used in a *post facto* image reconstruction of the original scene. Typically such masks are taken to be vector of components randomly and uniformly distributed, but here we use masks that represent specific Fourier components which we have found to be more efficient for our purposes. We denote the mask functions by $r_{k,w}(\mathbf{x})$, $k = 1, ..., K$. The projective measurements are simply the overlap integrals of these mask functions with the image,

$$m_{k,w}^{(j,\ell,\alpha)} = \int g_w^{(j,\ell,\alpha)}(\mathbf{x}) r_{k,w}(\mathbf{x}) d\mathbf{x} + n_k, \tag{6}$$

where the additive noise n_k for the kth measurement is assumed to be distributed according to a Gaussian PDF with zero mean and variance σ_N^2.

The essential metric of performance in the statistical IT description [1] we now present is the ratio of the mutual information, which is the information successfully passed by the measurement system, to the entropy, which represents the maximum information presented to the system. This ratio can be calculated for each parameter of interest as a way of assessing the overall system performance. We demonstrate our IT formalism first for the special case where the boundary location α is known *a priori*. In this simplified problem, the only variable of interest is the flux vector, $\mathbf{I} = (i^{(1,j)}, i^{(2,l)})$, whose two components denote the flux density of materials 1 and 2, respectively. We shall take the statistical distribution of material pairs to follow a uniform PDF,

$$p(\mathbf{I}) = \frac{1}{N_{pairs}}, \tag{7}$$

where the total pair combinations of materials on the target is $N_{pairs} = N_m(N_m-1)/2$ where N_m is the total number of material pairs in the database (of spectral signatures) of known materials. The PDF for the additive noise vector, $\mathbf{n}$, is of the Gaussian form

$$p(\mathbf{n}) = \frac{1}{(2\pi\sigma_N^2)^{K/2}} \exp\left[-\mathbf{n}^T\mathbf{n}/(2\sigma_N^2)\right], \tag{8}$$

where the subscript T on a vector or matrix represents its transpose and a matrix-product operation is understood whenever two vectors or matrices occur next to each other.

For a discrete-valued random variable, the above self-average must be understood as being calculated over the discrete probability distribution $P_X(x_i)$ via Eq. (1).When the discrete probability distribution is derived from a continuous PDF by integrating the latter over successive discreteness intervals, or histogram bins, each of width Δ_X that is very small compared to the scale over which the PDF varies, then, since $P_X(x_i) \approx p_X(x_i)\Delta_X$, we may relate the differential and discrete entropies as

$$H(X) \approx h(X) - \ln \Delta_X. \tag{9}$$

This relation breaks down when the discreteness interval Δ_X, which we may also call the precision to which the continuous variable X is "measured," or simply the X-precision, is no longer small compared to the width of the PDF. In this paper we shall assume that the relation (9) is always valid and therefore we can utilize it as a way of computing the discrete entropy from its corresponding differential entropy.

Since MI is a difference of two entropies for the same variable, it is well approximated according to relation (9) by the difference of the corresponding differential entropies from which the $\ln \delta X$ terms drop out. As we indicated earlier, the ratio, $\mathcal{R} \equiv I(X;Y)/H(X)$, may be regarded as a measure of the fidelity with which information about X can be extracted from the measurement Y.

We first consider a sequential set of measurements performed in a single wavelength band w. The conditional PDF for the measurement vector $\mathbf{m}$, given $\mathbf{I}$ and α, is seen to be simply the K-dimensional hyperspherical Gaussian PDF of the IID noise vector,

$$p_w(\mathbf{m}|\mathbf{I},\alpha) = \frac{1}{(2\pi\sigma_N^2)^{K/2}} \prod_{k=1}^{K} \exp\left[- \parallel \mathbf{m} - i_w^{(1,j)}\mathbf{p}_w^{(1,\alpha)} - i_w^{(2,l)}\mathbf{p}_w^{(2,\alpha)} \parallel^2 /(2\sigma_N^2)\right] \tag{10}$$

where the subscript w denotes the wave band and a pair of double vertical bars enclosing a vector $\mathbf{v}$ denotes its Euclidean norm, $\parallel v \parallel \equiv \sqrt{\mathbf{v^T v}}$. Because of its product form of the Gaussian PDF, its differential entropy is simply

$$h(\mathbf{m}|\mathbf{I},\alpha) = (K/2)\ln\left(2\pi e\sigma_N^2\right). \tag{11}$$

We assume a uniform distribution on the boundary parameters and then use Bayes rule to compute the conditional PDFs, $p_w(\mathbf{m}|\alpha)$ and $p_w(\mathbf{m}|\mathbf{I})$, and then via Eq. (2) compute the mutual information in the measurements about the surface materials of the two components and the parameters that describe the common boundary.

The measurement PDF in Eq. (10) is for a single waveband. We compute the measurement PDF for spatially compressive sensing system, operating at multiple wavelength bands in the region of the spectrum starting at $0.4\mu m$ and ending at $2.5\mu m$, as the W-fold product of Eq. (10),

$$p(\mathbf{m}|\mathbf{I},\alpha) = \prod_{w=1}^{W} p_w(\mathbf{m}|\mathbf{I},\alpha). \tag{12}$$

As in the case of a single band measurement, we compute both, $p(\mathbf{m}|\alpha)$ and $p_w(\mathbf{m}|\mathbf{I})$, and the compute the mutual information in the measurements performed in each wavelength band.

Computation of the statistical entropies for Eq. (10) and Eq. (12) relies on an implementation of a Markov Chain Monte Carlo (MCMC) algorithm that obtains a sequence of random samples from the probability distribution of interest. This sequence is used to approximate the distribution and compute the statistical entropies. The outline of our approach follows. We start with a random measurement value $\mathbf{m}$ and draw the next measurement, $\mathbf{m}^*$, based on a proposal distribution q centered on $\mathbf{m}$. We evaluate the ratio,

$$\gamma = \min\left(\frac{p(\mathbf{m}^*)q(\mathbf{m}_t^*,\mathbf{m}_{t-1})}{p(\mathbf{m}_{t-1})q(\mathbf{m}_{t-1},\mathbf{m}^*)}, 1\right) \tag{13}$$

and accept the proposed measurement value with probability γ. This implementation is referred to as the Metropolis-Hastings algorithm[3, 4] and it enables us to compute statistical entropies efficiently in high dimensions.

Important factors that will reduce the amount of information in the compressive measurements include the amount of blur, the noise and the choice of measurement mask. The image blur in Eq. (4) acts as a spatial filter that reduces information about the scene, particular at high spatial frequencies. The amount of filtering is greater at longer wavelengths than at shorter wavelengths. Reducing the effects of noise on the measurement can be accomplished by integrating over many wavelength channels; however, due to the wavelength dependence of the blur the integrated image becomes a sum of images with differing amounts of blur. The ability of the measurement mask to capture information about the parameters of interest in the presence of noise and this *integrated* blur can be evaluated by computing the mutual information for different choices of masks.

3 Results

We demonstrate this analysis on the problem in Fig. 2 for a spatially compressive sensing system operating in multiple wavelength bands. In each wavelength band a sequential set of measurements is performed using the DC mask, the sine/cosine for the lowest spatial frequency and the sine/cosine for the next lowest spatial frequency for a total of 5 measurements. The masks for the first three measurements are shown in Fig. 3. Our results are shown in Fig. 4.

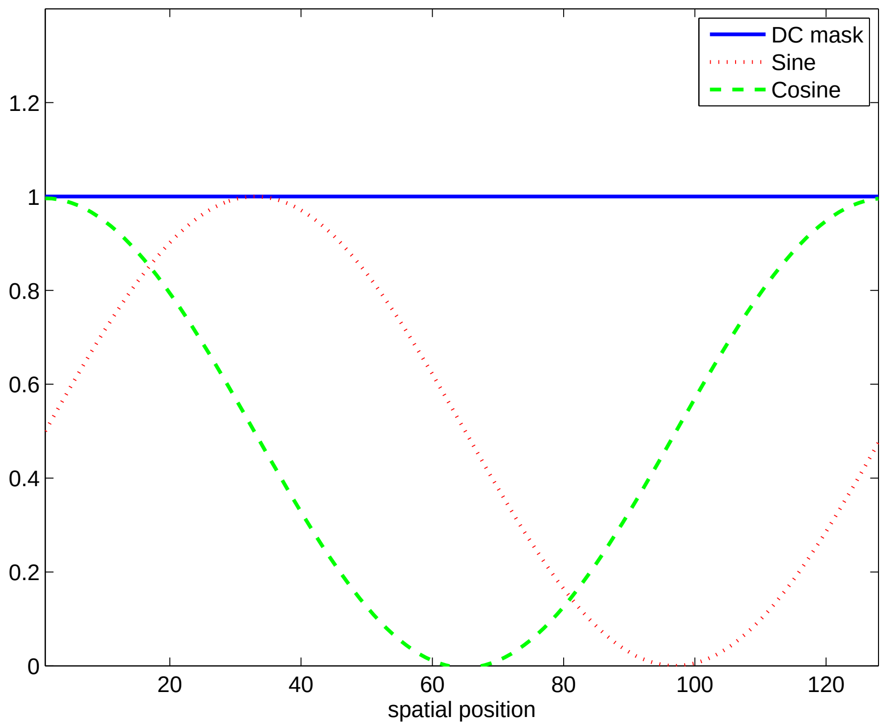


Figure 3: **Plotted are the DC (zero Fourier frequency) and the cosine and sine components of the lowest spatial frequency**

Here the blue curve (square symbols) represents the MI recovered using a single DC measurement mask at a different number of wavebands. In this scenario all the information about the materials comes from simply measuring the total flux in each band. Clearly, in the limit of no noise such an approach would recover all the information about the materials, which in our case is $\log_2(N_{pairs})$= 5.81 bits (here we have expressed the recovered MI as a fraction of the entropy $H(\mathbf{I})$). As more wavebands used the MI recovered increases due to the wavelength-dependent reflectance properties of the materials, as they are more easily identified. However, beyond 4 wavebands the MI recovered begins to decrease due to the splitting of photons amongst the multiple wave-bands. Here, our information-based analysis shows the trade-off point when using a mutli-wavelength compressive sensing approach.

In addition to the DC measurement we now included spatial measurements using the cosine and sine components of the lowest spatial frequency, i.e. we perform 3 measurements. One can see the added information by including a spatial measurement with the DC measurement. Like the DC measurement alone, we also see a decrease in information as more wave bands are introduced; however this decrease is mitigated by the additional measurements. We next include the second lowest spatial frequency for at total 5 measurements. One can now quantify the additional information provided by this measurement, and see further mitigation of the information loss at 10 wavebands.

3.1 Conclusions

We have analyzed the problem of SOI from spatially compressive sensing measurements from an information theoretic viewpoint . We have shown that by computing MI one can assess the additional information acquired by using multiple wave bands. It was shown that increasing the number of wavebands results in an increase in MI, which is due to the wavelength-dependent reflectance properties of the materials. As more wavebands are introduced however, the MI begins to decrease due to the splitting of photons amongst the wavebands. The MI allows one to quantify this trade-off between SNR and enhanced material discrimination.

The above results rely heavily on the Metropolis-Hastings algorithm, while the result shown in Fig. 4 for five measurements in 10 wave bands depends on computing the statistical entropies, i.e. Eq. (1), in a 50-dimensional parameter space. As the complexity of our problem increases, we will begin to encounter measurement probability functions that

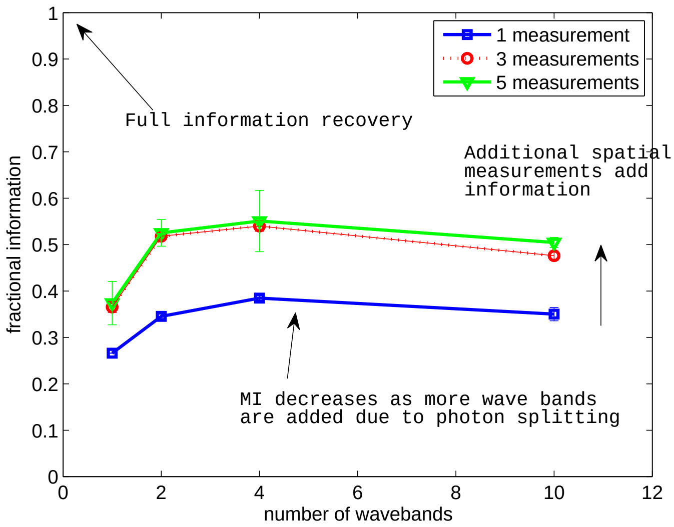


Figure 4: **The MI recovered about boundary parameters α and $\underline{I}$ using a spatially compressive sensing system operating at multiple wavelength bands**

can, depending on our local boundary parameterization and the reflectance properties of the materials chosen from our database, have multiple modes. An example is shown in Fig. 5. Our next steps will be the implementation of a Normal Kernel Coupler [5] to address entropy computations for similar cases. Eventually, our simulations will be extended to include the case of both a spatial and spectral compressive system [6] for more comprehensive results.

Acknowledgment

The work reported here was supported by the US Air Force Office of Scientific Research under award FA9550-09-1-0495.

References

[1] C. Shannon, The Bell System Technical Journal,Vol. 27, pp. 379-423, 623-656, July, October, 1948.

[2] T. Cover and J. Thomas, "Elements of Information Theory," Wiley (1991), Chap. 2. formula 6.631 (10), p. 718.

[3] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, "Equations of State Calculations by Fast Computing Machines," Journal of Chemical Physics, vol. 21 (6), pp. 1087 1092. (1953)

[4] W.K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," Biometrika, vol. 57 (1), pp. 97 109 (1970)

[5] G.R. Warnes, "The Normal Kernel Coupler: An adaptive Markov Chain Monte Carlo method for efficiently sampling from multi-modal distributions," Technical Report No. 395, Department of Statistics, University of Washington, Seattle, WA (2001)

[6] Ashwin Wagadarikar, Nikos Pitsianis, Xiaobai Sun, and David Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," Optics Express, 17 (8) (2009).

[7] E. Candés and M. Wakin, "An introduction to compressive sampling," IEEE Signal Processing Magazine, vol. 25, pp. 21 - 30 (2008).

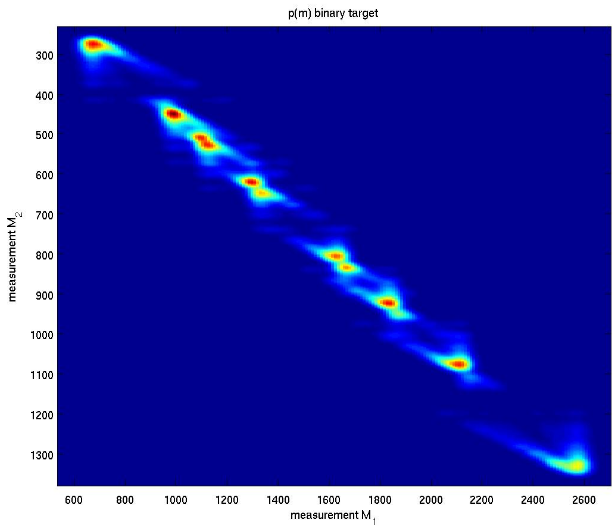


Figure 5: **An example of a 2-dimensional measurement probability function exhibiting multiple modes.**