

Scalable Information Optimal Compressive Target Recognition

Ronan Kerviche^a and Amit Ashok^{a,b}

^aCollege of Optical Sciences, University of Arizona, Tucson, AZ, USA

^bDepartment of Electrical and Computer Engineering, University of Arizona, Tucson, AZ, USA

ABSTRACT

We present a scalable information-optimal compressive imager optimized for the target classification task, discriminating between two target classes. Compressive projections are optimized using the Cauchy-Schwarz Mutual Information (CSMI) metric, which provides an upper-bound to the probability of error of target classification. The optimized measurements provide significant performance improvement relative to random and PCA secant projections. We validate the simulation performance of information-optimal compressive measurements with experimental data.

Keywords: Compressive Imaging, Classification, Cauchy-Schwarz Mutual Information, Target Recognition.

1. INTRODUCTION

A compressive imager exploits the relative sparsity/compressibility of natural scenes to acquire a small fraction of measurements relative to the number of isomorphic measurements acquired by a traditional imager.¹ In this work, we leverage the scalable parallel compressive imaging architecture reported in our prior work² for implementing the task of target recognition. The compressive projection/measurements for this specific task are optimized with the Cauchy-Schwarz mutual information (CSMI) measure that admits an analytical expression for Gaussian mixture model (GMM) of class conditional distributions for each target type. We demonstrate the resulting information-optimal compressive measurement design outperforms the random projections and other designed projections for the target recognition task.

2. LINEAR MEASUREMENT MODEL

We partition the scene into E non-overlapping blocks each of size N . For each scene block we make M compressive measurement that can be defined as:

$$\mathbf{g} = (I_E \otimes P)\mathbf{f} + \mathbf{w} \quad \text{or equivalently :} \quad G = PF + W, \quad (1)$$

where \mathbf{f} is the scene vector in \mathbb{R}^{EN} , \otimes denotes the Kronecker product, I_E is the identity matrix of dimension E , P is the $\mathbb{R}^{M \times N}$ projector matrix duplicated for each scene block, \mathbf{w} is a random vector in \mathbb{R}^{EM} denoting additive noise, and \mathbf{g} is the \mathbb{R}^{EM} vector of compressed measurements. In the following sections we will assume that the noise is additive white Gaussian (AWGN) with a variance of σ_n^2 , *i.e.* $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 I_{EM})$. We can also express this measurement model into a simple block-wrapped expression where F is a matrix in $\mathbb{R}^{N \times E}$, G and W are both $\mathbb{R}^{M \times E}$ matrices. Finally, as the compressive measurement operator P is implemented with passive optics, all the elements of the operator are less than (or equal to) unity. To incorporate the measurements constraints implemented by our compressive architecture we will enforce that all projections (*i.e.* rows of P) have equal exposures. This is equivalent to constraining all elements of P in the range $[-1/M; 1/M]$.



Figure 1. (Top row) Examples of mixture components (mean) comprising the training set. (Bottom row) Sample test scenes.

3. CAUCHY-SCHWARZ MUTUAL INFORMATION DESIGN METRIC

For the target recognition task, the ideal compressive measurement design objective function is the probability of error P_e , also referred to as the misclassification rate. However, this metric cannot be easily computed for any non-trivial measurement distributions. In Information theory, Kovalevskij's³ and Fano's⁴ inequalities respectively provide upper and lower bounds on P_e based on the Shannon mutual-information, however, here again it is not computationally tractable to evaluate such bounds for realistic measurement distributions such as mixture models. The numerical approximation of the bound integrals involved quickly becomes computationally intractable for large signal dimensions. In this study, we use instead the Cauchy-Schwarz Mutual Information (CSMI).⁵ If we note \mathbf{G} a continuous random vector taking values in \mathbb{R}^N and C a discrete random variable in the ensemble $\{1, \dots, N_c\}$, where N_c is the number of target classes, then CSMI between these variables is defined as the natural logarithm of the cosine between the joint distribution $p_{\mathbf{G},C}$ and the product of the marginals $p_{\mathbf{G}}p_C$:

$$I_{CS}(\mathbf{G}, C) = D_{CS}(p_{\mathbf{G},C}, p_{\mathbf{G}}p_C) = -\ln \left(\frac{\sum_{c \in \mathcal{C}} \int_{\mathbb{R}^N} p_{\mathbf{G},C}(\mathbf{g}, c) p_{\mathbf{G}}(\mathbf{g}) p_C(c) d\mathbf{g}}{\sqrt{\sum_{c \in \mathcal{C}} \int_{\mathbb{R}^N} p_{\mathbf{G},C}(\mathbf{g}, c)^2 d\mathbf{g} \sum_{c \in \mathcal{C}} p_C(c)^2 \int_{\mathbb{R}^N} p_{\mathbf{G}}(\mathbf{g})^2 d\mathbf{g}}} \right). \quad (2)$$

Where D_{CS} denotes the Cauchy-Schwarz divergence. It ranges from 0, when the variables are independent, to $\ln(N_c)/2$. If we consider a simple example where the likelihoods of the classes are multivariate Gaussians, *i.e.* $p(\mathbf{g}|c=1) = \mathcal{N}(\mathbf{g}|\boldsymbol{\mu}_1, \Sigma_1)$ and $p(\mathbf{g}|c=2) = \mathcal{N}(\mathbf{g}|\boldsymbol{\mu}_2, \Sigma_2)$ in \mathbb{R}^N , and each class is equiprobable, then we can write the analytical expressions for both Bhattacharyya coefficient B_c , which will give us an upper-bound to P_e , and CSMI :

$$B_c = \int_{\mathbb{R}^N} \sqrt{\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_1, \Sigma_1) \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_2, \Sigma_2)} d\mathbf{x} = (8\pi)^{\frac{N}{2}} |\Sigma_1 \Sigma_2|^{\frac{1}{4}} \mathcal{N}(\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2, 2(\Sigma_1 + \Sigma_2)), \quad (3)$$

$$I_{CS} = \frac{1}{2} \ln(2) - \frac{1}{2} \ln \left(1 + 2 \frac{\mathcal{N}(\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2, \Sigma_1 + \Sigma_2)}{(4\pi)^{-\frac{N}{2}} (|\Sigma_1|^{-\frac{1}{2}} + |\Sigma_2|^{-\frac{1}{2}})} \right). \quad (4)$$

If we now consider two classes, each specified by a mixture of Gaussian composed of S_1, S_2 equally weighted components, where each component has the same covariance matrix : $p(\mathbf{g}|c) = \sum_{k=1}^{S_c} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma)/S_c$, then we can express CSMI as an upper-bound to the probability of error :

$$P_e \leq \frac{\max\{S_1, S_2\}}{2} \sqrt{\exp(\ln(2) - 2I_{CS}) - 1}. \quad (5)$$

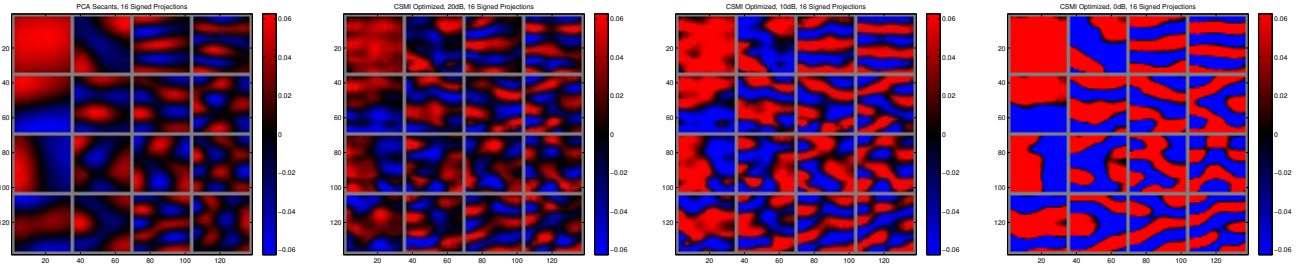


Figure 2. Optimization of the 16 32×32 projection patterns, from left to right : secant projections used as starting point, CSMI optimized projections for 20dB, 10dB and 0dB measurement SNR.

4. TARGET STATISTICAL MODEL

We employ a Gaussian mixture model (GMM) to learn the class conditional distribution for each target class, which is used both for compressive measurement optimization and the actual target classification from compressed measurements. A GMM can be expressed as follows:

$$p(\mathbf{f}|c) = \frac{1}{S_c} \sum_{k=1}^{S_c} \mathcal{N}(\mathbf{f}|\mathbf{r}_{c,k}, \sigma^2 I_{EN}), \quad (6)$$

where S_c is the number of mixture components for a given target class c and $\mathbf{r}_{c,k}$ is mean of the k^{th} mixture component and $\sigma^2 I_{EN}$ is a spherical covariance shared by all mixture components. We construct the class conditional distribution for each target class by finely sampling the corresponding target manifold and coarsely sampling the orthogonal background manifold. First, we generate a small set of blurred backgrounds images, which we refer to as *centroids*, by randomly sampling high resolution images acquired from scale models of target and background. On top of each of these centroids, we place templates of two different targets: a jeep and an Armored Project Vehicle (APV) by uniformly sampling the five degrees of freedom: two for transverse translation, one for scale/zoom, two for in-plane and out-of-plane rotations. In order to limit the complexity of the model, we limit the images to 2×2 blocks of 32×32 pixels. The targets can be placed in any position within this field of view except for a 10% margin at the edge of the field of view. The target scale is varied between 70% and 90% of the full image size (64×64 pixels). We allow in-plane rotation from -30° to 30° from the horizon. The out-of-plane rotation was captured by acquiring pictures of the scale target models under several viewing angles over the full 0° to 360° range. All the targets are slightly blurred so that their power spectral density is on the same scale as the background.

For the purpose of training the mixture model as well as testing the system performance of the compressive imager we generate two sets of (random) samples of targets embedded in the background. The targets are placed on the background by choosing random parameter values for their position (translations), rotations and scale, each within the limits described earlier. Based on the aforementioned model construction approach, we generate a mixture model for each of the two target classes comprised of 1.5 million equally weighted mixture components embedded in 4096 (64×64) dimensions. To optimize the mixture variance σ^2 , we test this mixture model against the first test set with one thousand images and record the variance value that yields the smallest probability of error obtained with the Maximum-A-Posteriori (MAP) classifier.

Based on these two target GMM models, we obtain the following expression for the mixture model describing the noise corrupted compressive measurements derived from a projection operator P of rank M ,

$$p(\mathbf{g}|c) = \frac{1}{\sqrt{2\pi}^{EM} S_c} |\sigma^2 P P^\top + \sigma_n^2 I_M|^{-\frac{E}{2}} \sum_{k=1}^{S_c} \exp\left(-\frac{1}{4} \text{Tr}\left[(G - S_{c,k})^\top (P^\top (\sigma^2 P P^\top + \sigma_n^2 I_M)^{-1} P) (G - S_{c,k})\right]\right) \quad (7)$$

where $S_{c,k}$ represents the projected k^{th} mixture component mean for the class c : $S_{c,k} = P R_{c,k}$. From this mixture distribution of compressive measurements we can derive an analytical expression for CSMI that can be

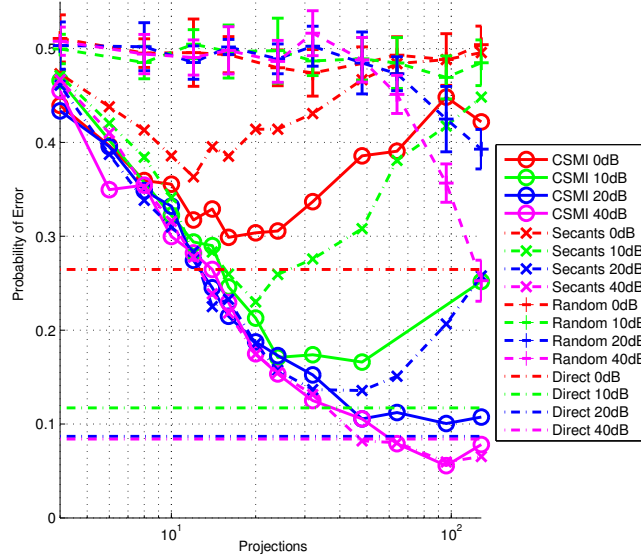


Figure 3. Performance of the CSMI optimized operators, compared to both PCA Secant and random (Gaussian) operators and conventional (direct) imager, for various number of projections (M) and measurement SNR.

expressed in terms of three different potential terms (i.e. $V_{\text{In}}, V_{\text{All}}, V_{\text{Between}}$) :

$$I_{CS}(\mathbf{G}, C) = \frac{1}{2} \ln(V_{\text{In}}) + \frac{1}{2} \ln(V_{\text{All}}) - \ln(V_{\text{Between}}), \quad (8)$$

$$V_{\text{In}} = \frac{1}{N_c^2} \sum_{p=1}^{N_c} \frac{1}{S_p^2} \sum_{i=1}^{S_p} \sum_{j=1}^{S_p} \mathcal{N}(\mathbf{s}_{p,i} | \mathbf{s}_{p,j}, 2I_E \otimes (\sigma^2 PP^\top + \sigma_n^2 I_{EM})), \quad (9)$$

$$V_{\text{All}} = \frac{1}{N_c^3} \sum_{p=1}^{N_c} \sum_{i=1}^{S_p} \sum_{q=1}^{N_c} \sum_{j=1}^{S_q} \frac{1}{S_p S_q} \mathcal{N}(\mathbf{s}_{p,i} | \mathbf{s}_{q,j}, 2I_E \otimes (\sigma^2 PP^\top + \sigma_n^2 I_{EM})), \quad (10)$$

$$V_{\text{Between}} = \frac{1}{N_c^3} \sum_{p=1}^{N_c} \sum_{i=1}^{S_p} \sum_{q=1}^{N_c} \sum_{j=1}^{S_q} \frac{1}{S_p S_q} \mathcal{N}(\mathbf{s}_{p,i} | \mathbf{s}_{q,j}, 2I_E \otimes (\sigma^2 PP^\top + \sigma_n^2 I_{EM})). \quad (11)$$

The number of interactions in this expression, i.e. terms within each potential, scales quadratically with the number of components S_p/S_q in each mixture. However, we observe that most of the Gaussians comprising each potential term are well separated, even after projection P i.e. in compressed measurement space. We can therefore limit the sum within each potential to only a few neighboring components and thereby reducing the computational complexity by nearly two orders of magnitude. The gradient of CSMI with respect to the operator P , can now be expressed as:

$$\nabla I_{CS}|_P = \frac{\nabla V_{\text{In}}|_P}{2V_{\text{In}}} + \frac{\nabla V_{\text{All}}|_P}{2V_{\text{All}}} - \frac{\nabla V_{\text{Between}}|_P}{V_{\text{Between}}}, \quad (12)$$

where the gradient of the potentials are dependent on the gradient of the multivariate Gaussians, with respect to the compressive measurement operator P that we wish to optimize:

$$\nabla \ln[\mathcal{N}(\mathbf{g}_1 | \mathbf{g}_2, 2I_e \otimes (\sigma^2 PP^\top + \sigma_n^2 I_M))] |_P = -E\sigma^2 \Sigma^{-1} P - \frac{1}{2} (\Sigma^{-1} P \Delta \Delta^\top - \sigma^2 \Sigma^{-1} P \Delta \Delta^\top P^\top \Sigma^{-1} P), \quad (13)$$

with $\Sigma = \sigma^2 PP^\top + \sigma_n^2 I_M$ and $\Delta = G_1 - G_2$, the difference of the block-wrapped components mean. Finally, the constrained compressive measurement (P) optimization problem can be stated as :

$$P_{\text{optimal}} = \underset{P}{\operatorname{argmax}} \{I_{CS}(P)\} \quad \text{s.t.} \quad |P_{i,j}| \leq \frac{1}{M} \quad (14)$$

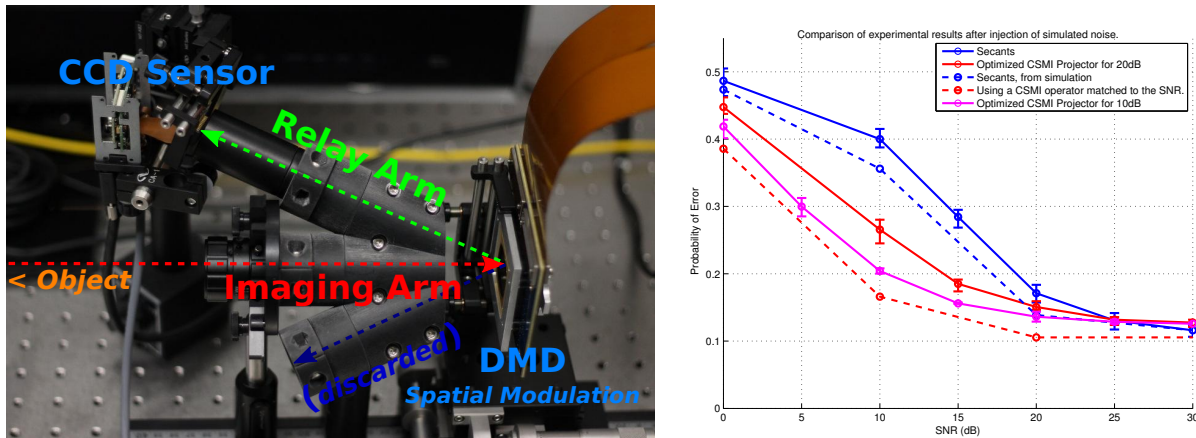


Figure 4. (Left) Scalable compressive imager prototype. (Right) Target classification performance of PCA Secant and CSMI optimized projection using simulation (dashed) and experimental (solid) data.

5. SIMULATION AND EXPERIMENTAL RESULTS

We optimize compressive measurement operator P , operating on 32×32 size block, for various compression ratio using PCA Secant projections as initial point.⁶ The PCA Secant projections are generated using the same training data used to derive the GMM target class models. In Figure 2, we show the results of the compressive measurement operator (P) optimization for various Signal to Noise Ratio (SNR) and the PCA Secant projections used as the starting point. As expected in the low SNR regime, the resulting optimized compressive operators tend towards nearly binary transmission patterns that maximize the photon efficiently to generate a meaningful signal. As the measurement SNR increases, the optimized operator progressively approach the PCA Secant design indicating their optimality in the high SNR regime. Using the MAP classifier, defined below, we quantify the target classification performance of various compressed measurement operators:

$$\hat{c}(\mathbf{g}) = \underset{C}{\operatorname{argmax}} \{p(c|\mathbf{g})\}. \quad (15)$$

The probability of target misclassification (error) rate P_e is obtained by testing a set of two thousand samples that are distinct from the training data. The performances of random (Gaussian) projections, PCA Secant and CSMI optimized operators are shown in Figure 3 as a function of number of measurements/projections (M). We observe that the CSMI optimized operator provides a significant improvement over the random and PCA Secant operators, especially in the low SNR regime. Here, for the task of target classification we are able to obtain measurement compression ratios ranging from $21\times$ to $42\times$. Note that each compressive measurement operator reaches its respective minima at different number of projections for a given SNR value. This is expected due to the fixed photon budget: using more measurements results in fewer photons per measurement, thus eventually degrading the performance due to decreasing measurement SNR.

To validate these simulation results, we perform experimental measurements with our programmable compressive imager testbed, shown in Figure 4, described in our prior work.² We acquire $M=48$ compressive measurements (i.e. $21\times$ compression ratio) at a SNR exceeding 30dB and we numerically add white gaussian noise to simulate lower SNR operating point. The target recognition performance of compressive measurements using simulation and experimental data in shown in Figure 4. Overall, we find good agreement between simulation and experiment data.

6. CONCLUSION

We have demonstrated that projection operators optimized with the CSMI metric can improve the target recognition performance significantly relative to random and PCA Secant projections. This performance improvement was validated with experimental data.

ACKNOWLEDGMENTS

Authors gratefully acknowledge support from Army STTR (W911NF-13-C-0066) and Arizona TRIF.

REFERENCES

- [1] D. Donoho, "Compressed sensing", in *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289-1306, 2006.
- [2] R. Kerviche, N. Zhu, and A. Ashok, "Information-optimal Scalable Compressive Imaging System," in *Classical Optics 2014*, OSA Technical Digest (online) (Optical Society of America, 2014), paper CM2D.2.
- [3] V.A. Kovalevskij, "The problem of character recognition from the point of view of mathematical statistics," *Character Readers and Pattern Recognition*, pp. 3-30, Spartan: New York, 1968. Russian edition 1965.
- [4] R.M. Fano, "Transmission of Information: A Statistical Theory of Communication," MIT: New York, 1961.
- [5] J. C. Principe, "Information Theoretic Learning, Renyi's Entropy and Kernel Perspectives," Springer 2010.
- [6] Y. Li, C. Hegde, A. C. Sankaranarayanan, R. Baraniuk and K. F. Kelly, "Compressive image acquisition and classification via secant projections," *Journal of Optics*, Volume 17(6), 2015.