MULTISCALE RANDOM PROJECTIONS FOR COMPRESSIVE CLASSIFICATION

Marco F. Duarte,[r] Mark A. Davenport,[r] Michael B. Wakin,[c] Jason N. Laska[r]
Dharmpal Takhar,[r] Kevin F. Kelly,[r] Richard G. Baraniuk[r]

[r] Department of Electrical and Computer Engineering, Rice University
[c] Department of Applied and Computational Mathematics, California Institute of Technology

ABSTRACT

We propose a framework for exploiting dimension-reducing random projections in detection and classification problems. Our approach is based on the generalized likelihood ratio test; in the case of image classification, it exploits the fact that a set of images of a fixed scene under varying articulation parameters forms a low-dimensional, nonlinear manifold. Exploiting recent results showing that random projections stably embed a smooth manifold in a lower-dimensional space, we develop the multiscale smashed filter as a compressive analog of the familiar matched filter classifier. In a practical target classification problem using a single-pixel camera that directly acquires compressive image projections, we achieve high classification rates using many fewer measurements than the dimensionality of the images.

Index Terms— Data Compression, Image Coding, Image Classification, Object Recognition

1. INTRODUCTION

Compressive sensing (CS) systems acquire and reconstruct compressible signals from a small number of non-adaptive linear random measurements by combining the steps of sampling and compression [1, 2]. CS enables the design of new kinds of compressive imaging systems, including a "single-pixel" camera with some attractive features, including simplicity, low power consumption, universality, robustness, and scalability [3]. Since it relies on a single photon detector, the single-pixel camera can be adapted to image in situations where conventional CCD and CMOS imagers are blind.

In many data acquisition/processing applications we are not interested in obtaining a precise reconstruction, but rather are only interested in making some kind of detection or classification decision. For instance, in target classification, we simply wish to identify the class to which an image belongs out of several possibilities. In this paper, we propose a new framework for *compressive classification* for such situations that bypasses the reconstruction and makes a decision based solely on random measurements. We pay particular attention to generalizing the matched filter to compressive measurements using the generalized likelihood ratio test (GLRT) and the fact that the set of images of a fixed scene from different camera positions forms a low-dimensional, nonlinear manifold. Exploiting recent results on the multiscale structure of image appearance manifolds (IAMs) and random projections of manifolds [4, 5], we design a multiscale pseudo-random measurement scheme and a novel classification algorithm—the *multiscale smashed filter*—that can be viewed as a generalization of the classical matched filter to more challenging manifold settings. In certain situations, the smashed filter achieves high classification accuracy using many fewer random measurements than would be needed to reconstruct the image under view. This work builds on the preliminary investigations of the smashed filter in [6] and other algorithms for compressive classification in [7–9].

This paper is organized as follows. Section 2 provides an overview of both the theoretical foundation of CS as well as a brief description of the single-pixel camera. Section 3 develops compressive classification and the smashed filter, and Section 4 presents preliminary experimental results. Section 5 concludes with a brief discussion of our results and directions for future work.

2. COMPRESSIVE SENSING

CS builds upon the fact that signals, images, and other data often contain some type of structure that enables intelligent representation and processing. As an example, many signals have a *sparse* representation in terms of some basis Ψ. We say that a signal $\mathbf{x} \in \mathbb{R}^N$ is K-sparse if it can be represented as $\mathbf{x} = \Psi\boldsymbol{\alpha}$ where the vector $\boldsymbol{\alpha} \in \mathbb{R}^N$ has only $K \ll N$ significant coefficients. The surprising core result of CS is that a length-N signal that is K-sparse can be recovered exactly from a nonadaptive linear projection of the signal onto a random $O(K \log(N/K))$-dimensional basis [1, 2]. Thus we can directly acquire the signal in a compressed form using random measurements.

Specifically, we encode $M = O(K \log(N/K))$ inner products of the signal with a set of random vectors. In matrix notation, we measure

$$\mathbf{y} = \Phi\mathbf{x}, \tag{1}$$

with $\mathbf{y}$ an $M \times 1$ column vector and the measurement matrix Φ an $M \times N$ random matrix. Since $M < N$, recovery of
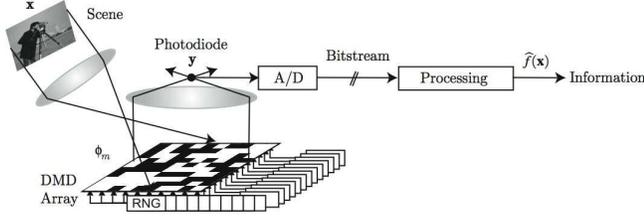
Fig. 1. *Single-pixel compressive camera block diagram. Incident lightfield (corresponding to the desired image x) is reflected off a digital micromirror device (DMD) array whose mirror orientations are modulated in the pseudo-random pattern ϕ_m supplied by a random number generator. Each different mirror pattern produces a voltage at the single photodiode that corresponds to one measurement $y(m)$.*

the signal **x** from the measurements **y** is ill-posed in general; however the additional assumption of signal *sparsity* makes recovery possible and practical [1, 2].

Imaging is a particularly compelling application of CS. Our single-pixel camera [3] employs a Texas Instruments digital micromirror device (DMD), which consists of an array of electrostatically actuated micromirrors. The camera focuses the desired image on the DMD plane; the light reflecting from a random subset of the mirrors is then focused onto a single photodiode, which measures the value of the inner product between the image and the random binary $\{0, 1\}$ pattern displayed on the DMD, thus yielding a measurement $y(m)$ for that pattern. By switching between M different random patterns, we can collect the entire measurement vector **y**.

This system directly acquires a reduced set of M random projections of an N-pixel image **x** without ever acquiring the N pixel values. Furthermore, the camera is *progressive*, meaning that better quality images can be obtained by taking more measurements. Other desirable features include the single detector (potentially enabling imaging at wavelengths too difficult or expensive to measure using sensor arrays), and the availability of universal measurement bases (since random matrices are incoherent with arbitrary sparsity bases).

In many applications, however, image acquisition is performed for purposes other than reconstruction. For example, in vision and surveillance settings images are processed to extract different kinds of information from the observed scene, varying from the presence or absence of a target object to a parametrization of the articulation of objects in view. It is now known that random CS measurements can also directly capture the necessary information from the image to perform such tasks [6–9]. Thus, the single-pixel camera and other CS acquisition devices can be applied to these problem without modification. In the next section, we develop a framework for target classification from compressive imaging measurements.

3. COMPRESSIVE CLASSIFICATION

In this section, we develop a new classification algorithm that uses compressive multiscale measurements to exploit the low-dimensional manifold structures inherent in target recognition applications. We learn this manifold structure from training

data, which correspond to a sampling of points from each manifold. The manifold structure enables us both to reduce the dimensionality of the training data through random measurements and to limit the amount of training data required to perform the classification.

3.1. Generalized Likelihood Ratio Test

We first review standard GLRT classification framework. Consider P possible image classes and define the hypothesis $\mathcal{H}_i$ to be that the observed image $\mathbf{x} \in \mathbb{R}^N$ belongs to class $\mathcal{C}_i$ for $i = 1, \ldots, P$. For each class $\mathcal{C}_i$, an element $\mathbf{x} \in \mathcal{C}_i$ can be parameterized by a unique K-dimensional parameter vector $\boldsymbol{\theta}_i \in \Theta_i$, i.e., $\mathbf{x} = f_i(\boldsymbol{\theta}_i)$ for some f_i; an example parameter is the pose of the object in the scene (translation, rotation, etc.). If the mapping f_i is well-behaved, then the collection of signals $\{f_i(\boldsymbol{\theta}_i) : \boldsymbol{\theta}_i \in \Theta_i\}$ forms a K-dimensional manifold embedded in the ambient image space $\mathbb{R}^N$.

First, assume that we take noisy measurements $\mathbf{y} = \mathbf{x} + \boldsymbol{\omega}$ of **x**, giving us the distribution $p(\mathbf{y}|\boldsymbol{\theta}_i, \mathcal{H}_i)$ for the measured signal **y** under hypothesis $\mathcal{H}_i$ and parameters $\boldsymbol{\theta}_i$. The GLRT classifier is given by

$$\mathcal{C}(\mathbf{y}) = \arg\max_{i=1,\ldots,P} p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_i, \mathcal{H}_i), \tag{2}$$

where

$$\widehat{\boldsymbol{\theta}}_i = \arg\max_{\theta \in \Theta_i} p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}_i) \tag{3}$$

denotes the maximum likelihood estimate (MLE) of the parameters $\boldsymbol{\theta}_i$ under hypothesis $\mathcal{H}_i$. Under an additive white Gaussian noise (AWGN) model for $\boldsymbol{\omega}$, the likelihood for each hypothesis $\mathcal{H}_i$ becomes

$$p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_i, \mathcal{H}_i) \propto \frac{1}{\|\mathbf{y} - f_i(\widehat{\boldsymbol{\theta}}_i)\|_2^2}, \tag{4}$$

meaning that after estimates for the parameters are obtained for each class, the GLRT reduces to nearest-neighbor classification among the available hypotheses.

3.2. Manifold Parameter Estimation

In order to implement the GLRT, we first need to obtain estimates of the parameter vectors $\widehat{\boldsymbol{\theta}}_i$ from the noisy measurements **y** under each of the hypotheses. A natural approach to this problem is through nonlinear least-squares, in which we seek the value of $\boldsymbol{\theta}_i$ that minimizes the objective function

$$D(\boldsymbol{\theta}_i) = \|\mathbf{y} - f_i(\boldsymbol{\theta}_i)\|_2^2. \tag{5}$$

For differentiable $D(\boldsymbol{\theta})$, we can use Newton's method to obtain iterative estimates of the parameters as

$$\boldsymbol{\theta}_i^n = \boldsymbol{\theta}_i^{n-1} - [\mathbf{H}(\boldsymbol{\theta}_i^{n-1})]^{-1} J(\boldsymbol{\theta}_i^{n-1}) \tag{6}$$

for the nth iteration, with $\mathbf{J}(\boldsymbol{\theta}) = \vec{\nabla} D(\boldsymbol{\theta})$ (the gradient) and $\mathbf{H}(\boldsymbol{\theta})$ the Hessian matrix of D; with a good starting point,

the algorithm converges to the correct estimate. Note that the classical matched filter is an elegant method for minimizing (5) on the manifold consisting of all possible shifts of a signal. In essence, (6) generalizes it to a richer class of manifolds, while reducing the number of samples from the manifold required during the estimation process.

A number of challenges remain, however. First, in general, implementing the estimator (6) requires complete knowledge of the function f_i or the ability to evaluate $f_i(\boldsymbol{\theta})$ for all possible values of $\boldsymbol{\theta}$. In some practical settings this may not be possible, but fortunately a dense sampling of the parameter space Θ_i and a nearest neighbor (NN) estimation rule often yield acceptable performance, albeit with a potentially high computational cost. Second, the manifolds under consideration may not be differentiable, in which case we cannot directly apply (6) [4]. Third, it may be possible that random projections of the data could alter the manifold structure of our signals. Fortunately, while the last two challenges are more challenging that the first, we can overcome them using *multiscale* measurements.

3.3. Multiscale Measurements for Image Appearance Manifolds

In the case of interest—target classification—the classes $\mathcal{C}_i$ are IAMs, with each manifold corresponding to a different class of targets. The parameter vector $\boldsymbol{\theta}_i$ denotes the articulation parameters for the target, such as rotation, translation, angle of view, etc. The resulting parametric manifolds are *nonlinear*—since linear combinations of manifold elements are in general not contained in the manifold—and *non-differentiable*—due to sharp edges in the images caused by occlusions and to sharp features rotating in/out of view.

Previous research [4] has identified a multiscale structure to such manifolds that can be exploited through regularization to obtain differentiable manifolds at a nested set of scales. The regularization is achieved through the use of a nested set of regularization kernels $G_1, G_2, \ldots$, with the kernels of decreasing support. Now, instead of minimizing (5) directly using Newton's method, we minimize a sequence of objective functions

$$D_n(\boldsymbol{\theta}_i) = \|G_n \mathbf{y} - G_n f_i(\boldsymbol{\theta}_i)\|_2^2, \tag{7}$$

that use the corresponding regularization kernels.

3.4. Compressive Measurements for Smooth Manifolds

It was recently shown that most of the structure of a *smooth* manifold is preserved under a random lower-dimensional projection [5]. That is, for a K-dimensional manifold embedded in N-dimensional space, with high probability a random M-dimensional projection is invertible—and thus preserves the manifold structure—provided that $M > CK \log(N)$ for some constant C that depends on the smoothness of the manifold. Thus, instead of performing parameter estimation directly on the signal $\mathbf{y} = \mathbf{x} + \boldsymbol{\omega}$, we can choose to observe only a lower dimensional, randomly projected version

$\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\omega}$, where Φ is an $M \times N$ measurement matrix with independent, randomly distributed entries. Accordingly, we update the objective function (5) to

$$D^C(\boldsymbol{\theta}_i) = \|\mathbf{y} - \Phi f_i(\boldsymbol{\theta}_i)\|_2^2. \tag{8}$$

When Gaussian random measurements are used, this is equivalent to employing different colored Gaussian random measurements at each iteration; see [10] for more details. Moreover, the dimensionality reduction affords savings in computational complexity and storage requirements of the estimation and classification algorithms described earlier.

3.5. The Multiscale Smashed Filter

We are now in a position to overcome the challenges from the end of Section 3.2. In [6] we introduced the *smashed filter* as a classification algorithm for compressive measurements of signal classes that are low-dimensional manifolds. The smashed filter is inspired by the fact that random projections do not disturb the structure of smooth manifolds, as described above. However, as we have just observed, in our setting the manifolds might not be smooth. To address this problem, we exploit the multiscale structure of IAMs and combine multiscale measurements with random projections by smoothing the target class IAMs so that their projections preserve their geometry.

The *multiscale smashed filter* uses a measurement matrix of the form

$$\Phi = \begin{bmatrix} \Phi_1 G_1 \\ \vdots \\ \Phi_S G_S \end{bmatrix},$$

where Φ_n is an $M_n \times N$ matrix with randomly distributed entries and G_n is the regularization kernel for the nth scale. The resulting measurements can be partitioned into measurements for each of the regularized versions, i.e., $\mathbf{y}_n = \Phi_n G_n \mathbf{x} + \boldsymbol{\omega}_n$, which are used in sequential iterations of Newton's method by employing the corresponding objective functions

$$D_n^C(\boldsymbol{\theta}_i) = \|\mathbf{y}_n - \Phi_n G_n f_i(\boldsymbol{\theta}_i)\|_2^2. \tag{9}$$

This classification algorithm employs the compact and multiscale nature of the target manifolds to estimate the signal parameters under each class hypothesis, together with the GLRT/NN classification rule from Section 3.1.

3.6. Hallmarks of Compressive Classification

In addition to the computational and storage savings achieved by compressive classification, our methods share many of the advantages of CS reconstruction. In particular, random projections enable *universal* estimation and classification, in the sense that random projections preserve the structure of any low-dimensional signal class with high probability. Additionally, random projections are progressive in the sense that a larger number of projections translate into higher classification rates thanks to increased noise tolerance. Finally, random projects are scalable, since the number of projections required

| (a) Tank | (b) School Bus | (c) Truck |

Fig. 2. *Models for the classification experiments.*

Fig. 3. *Multiscale smashed filter performance results. (Top) Vehicle classification as a function of number of measurements for different numbers of initial estimates. (Bottom) Average rotation estimate error as a function of the number of initial estimates.*

for accurate classification depends only logarithmically on the image size and the number of classes [6].

4. EXPERIMENTAL RESULTS

We performed experiments to evaluate the multiscale smashed filter in a target classification setting using real data from the single-pixel camera described in Section 2. We define three classes, each for a different vehicle model: a T-72 tank, a school bus, and a truck. The unknown parameter in each vehicle class is the rotation angle. The models are shown in Figure 2. For each of the vehicles, multiscale measurements were taken using five different resolutions – from 8×8 to 128×128 pixels – with the same number of measurements taken at each resolution. Measurements were taken each $2°$ of rotation (180 angles total).

The measurements for each rotation/class combination were classified using a multiscale smashed filter. For each of the target classes, one of the sampled rotations was chosen at random as an initial estimate; this selection was repeated multiple times to overcome local minima in the objective function. The gradient of the manifold $f_i(\boldsymbol{\theta}_i)$ was estimated using consecutive points in the manifold sampling, including that of the current estimate. We then executed Newton's method using measurements at different resolutions at each iteration, proceeding from the coarsest to the finest scale. After the rotation was estimated under each hypothesis, NN classification was performed. We repeated the experiment 10,000 times for each testing point, with randomly selected starting points each time, and we varied the number of measurements taken from 5 to 60. The results are shown in Figure 3 and indicate that thanks to the low-dimensional structure of the underlying IAM, very few measurements are necessary to achieve high classification rates. Additionally, the use of multiple initial estimates enables the algorithm to overcome issues with local minima and achieve lower average error for the rotation estimates.

5. CONCLUSIONS

We have developed a new framework for compressive classification based on random measurements that is designed for signals, images, and other data that lie on low-dimensional manifolds. The multiscale smashed filter significantly reduces the amount of data that must be acquired or stored for attaining a given classifier performance. Avenues for future research include reducing the sensitivity of the algorithm to the initial parameter estimate and further testing on real data from the single-pixel camera and other CS data acquisition systems.
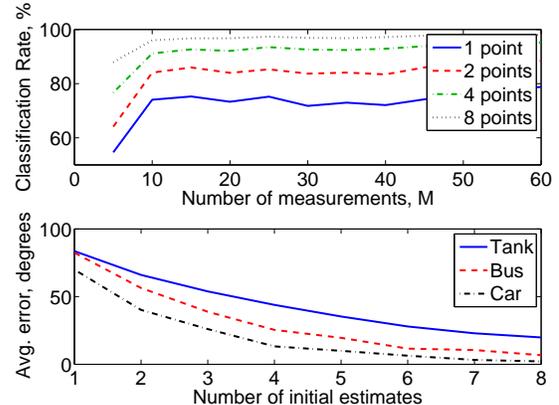
6. REFERENCES

[1] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Info. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, September 2006.

[3] D. Takhar, J. N. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. K. Kelly, and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," in *Proc. IS&T/SPIE Symposium on Electronic Imaging: Computational Imaging*, San Jose, CA, Jan. 2006, vol. 6065, pp. 43–52.

[4] M. B. Wakin and R. G. Baraniuk, "High-resolution navigation on non-differentiable image manifolds," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, PA, 2005, vol. 5, pp. V–1073–1076.

[5] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," 2006, Preprint.

[6] M. A. Davenport, M. F. Duarte, D. Takhar, J. N. Laska, K. K. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. IS&T/SPIE Symposium on Electronic Imaging: Computational Imaging*, San Jose, CA, Jan. 2007.

[7] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Sparse signal detection from incoherent projections," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 2006, vol. III, pp. 305–308.

[8] M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Detection and estimation with compressive measurements," Tech. Rep. TREE0610, Rice University ECE Department, 2006.

[9] J. Haupt, R. Castro, R. Nowak, G. Fudge, and A. Yeh, "Compressive sampling for signal classification," in *Proc. 40th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2006.

[10] Z. Wang, G. R. Arce, and J. L. Paredes, "Colored projections for compressed sensing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, HI, 2007.