The Smashed Filter
for Compressive Classification and Target Recognition

Mark A. Davenport,[r] Marco F. Duarte,[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

The theory of compressive sensing (CS) enables the reconstruction of a sparse or compressible image or signal from a small set of linear, non-adaptive (even random) projections. However, in many applications, including object and target recognition, we are ultimately interested in making a decision about an image rather than computing a reconstruction. We propose here a framework for *compressive classification* that operates directly on the compressive measurements without first reconstructing the image. We dub the resulting dimensionally reduced matched filter the *smashed filter*. The first part of the theory maps traditional maximum likelihood hypothesis testing into the compressive domain; we find that the number of measurements required for a given classification performance level does not depend on the sparsity or compressibility of the images but only on the noise level. The second part of the theory applies the generalized maximum likelihood method to deal with unknown transformations such as the translation, scale, or viewing angle of a target object. We exploit the fact the set of transformed images forms a low-dimensional, nonlinear manifold in the high-dimensional image space. We find that the number of measurements required for a given classification performance level grows linearly in the dimensionality of the manifold but only logarithmically in the number of pixels/samples and image classes. Using both simulations and measurements from a new single-pixel compressive camera, we demonstrate the effectiveness of the smashed filter for target classification using very few measurements.

Keywords: Compressive sensing, image classification, object recognition, smashed filter

1. INTRODUCTION

The theory of *compressive sensing* (CS) enables a sparse or compressible signal to be reconstructed from a small set of non-adaptive linear projections.[1,2] In some applications, this allows us to directly acquire a compressed representation of a signal, effectively combining the steps of sampling and compression. However, in many applications we are not interested in obtaining a precise reconstruction of the scene under view, 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 our image belongs out of several possibilities.

In this paper, we propose an algorithm to support a new theory of *compressive classification* that enjoys the same benefits as compressive sensing. Our approach, based on a generalized maximum likelihood classifier (GMLC), is applicable to a wide variety of signal classification problems. Focusing on the problem of image classification, we use the fact that the set of images of a fixed scene with different imaging parameters (translation, scale, view angle, illumination, etc.) forms a low-dimensional, nonlinear manifold in the high-dimensional ambient image space. Exploiting recent results on random projections of manifolds,[3] we design a pseudo-random measurement scheme and a new classification algorithm—the *smashed filter*—that can be viewed as a generalization of the classical matched filter to more challenging manifold settings. The smashed filter achieves high classification rates using only a small fraction of measurements compared to the dimensionality of the original images. This work builds on preliminary investigations of compressive classification.[4–7]

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 our prototype compressive imaging architecture. Section 3 provides background on conventional detection and estimation problems. Section 4 develops compressive classification and the smashed filter, and Section 5 presents preliminary experimental results using both simulations and real data from the single-pixel camera. Section 6 concludes with a brief discussion of our results and directions for future work.

2. BACKGROUND ON COMPRESSIVE SENSING

Compressive sensing (CS) builds upon a core tenet of signal processing and information theory: 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. In particular, we say that a signal $\mathbf{x} \in \mathbb{R}^N$ is K-sparse if it can be represented as $\mathbf{x} = \Psi\boldsymbol{\theta}$ where the vector $\boldsymbol{\theta} \in \mathbb{R}^N$ has only $K \ll N$ nonzero coefficients. We say that a signal is compressible if it can be closely approximated as K-sparse. The surprising result of CS is that a length-N signal that is K-sparse/compressible in some basis can be recovered exactly/approximately 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.

Specifically, rather than sampling the signal, we encode $M = O(K \log(N/K))$ inner products of the signal with a set of random vectors. In matrix notation, we take the *compressive measurements*

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

where $\mathbf{y}$ is an $M \times 1$ column vector, and Φ is an $M \times N$ random (but known) matrix. We will assume that Φ is an orthoprojector, i.e., that it has orthonormal rows. Since $M < N$, recovery of the signal $\mathbf{x}$ from the measurements $\mathbf{y}$ is ill-posed in general; however the additional assumption of signal sparsity or compressibility makes recovery both possible and practical.

Imaging is a particularly compelling application of CS. Our single-pixel camera[8] employs a Texas Instruments digital micromirror device (DMD), which consists of an array of N electrostatically actuated micromirrors. The camera focuses the light from the desired image onto the DMD with the mirrors set in a pseudorandom binary (0,1) pattern; the light reflecting from the mirrors set to 1 is then focused onto a single photodiode. The voltage at the photodiode is thus the inner product between the image and the random binary (0,1) pattern displayed on the DMD. Switching among
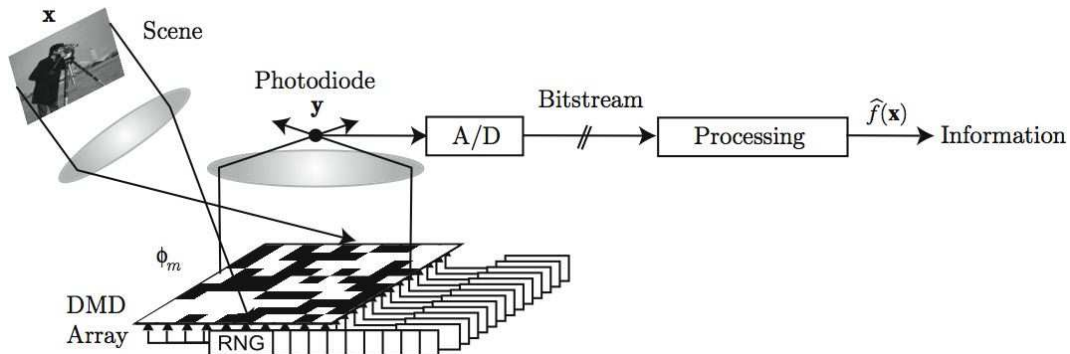
Figure 1. *Single-pixel compressive imaging camera block diagram. Incident lightfield (corresponding to the desired image $\mathbf{x}$) is reflected off a digital micromirror device (DMD) array whose mirror orientations are modulated in the pseudorandom pattern ϕ_m supplied by a random number generator. Each different mirror pattern produces a voltage at the single photodiode that corresponds to one measurement y_m.*

M different pseudorandom patterns collects a sufficient amount of information to reconstruct or approximate the N-pixel image.

This single-pixel camera enjoys a number of advantages over traditional imaging systems. It is *universal* in the sense that the random measurements do not depend on the basis Ψ in which the image is sparse or compressible. The measurement process is also *progressive* in that better quality images can be obtained by simply taking more measurements. Compressive measurements are *democratic* in that each measurement can be given equal priority and that the reconstruction quality depends only on how many measurements are received, not on the particular subset received; this is particularly useful in remote sensing and distributed sensing where communication may be unreliable. Additionally, the use of a single detector enables imaging at new wavelengths unaccessible or prohibitively expensive using current focal plane imaging arrays.

In many applications, however, image acquisition is performed for purposes other than compression and storage. For example, in vision and surveillance settings, images are processed to extract different amounts of information from the scene observed, varying from the presence or absence of a target object to a full parametrization of the articulations of the objects in view. It is now known that compressive measurements can capture the necessary information from the image to perform such tasks.[5–7] Thus, hardware introduced for compressive imaging can also be used for these applications without modification.

3. BACKGROUND ON CLASSIFICATION

In this section, we provide background on some conventional problems in statistical signal processing. In Section 4 we translate these to the compressive domain. While much of our discussion focuses on the specific case of image classification, our approach is applicable to general classification settings.

3.1. Maximum Likelihood Classification

We begin by examining the problem of signal classification using a *maximum likelihood classifier* (MLC). Suppose a signal $\mathbf{x} \in \mathbb{R}^N$ belongs to one of P classes $\mathcal{C}_i, i = 1, \ldots, P$. We let hypothesis

$\mathcal{H}_i$ signify that the signal $\mathbf{x}$ belongs to class $\mathcal{C}_i$, and for now we assume that class $\mathcal{C}_i$ contains a single known signal $\mathbf{s}_i$. We obtain noisy measurements of $\mathbf{x}$, as in $\mathbf{y} = \mathbf{x} + \boldsymbol{\omega} \in \mathbb{R}^N$, giving us a distribution $p(\mathbf{y}|\mathcal{H}_i)$ for the measured signal $\mathbf{y}$ under hypothesis $\mathcal{H}_i$. The MLC classifies according to which class has the maximum class-conditional likelihood

$$\mathcal{C}(\mathbf{y}) = \underset{i=1,\ldots,P}{\arg\max} \ p(\mathbf{y}|\mathcal{H}_i). \tag{2}$$

Under an additive white Gaussian noise (AWGN) model for $\boldsymbol{\omega}$ with variance σ, the probability distribution for the measured signal $\mathbf{y}$ under hypothesis $\mathcal{H}_i$ becomes

$$p(\mathbf{y}|\mathcal{H}_i) = \frac{1}{(2\pi\sigma)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma}\|\mathbf{y}-\mathbf{s}_i\|_2^2}.$$

In this case (2) reduces to a *nearest-neighbor* classifier among the available hypotheses. Note that in the case of classification between two equally likely classes, the MLC reduces to the more common likelihood ratio test.

3.2. Generalized Maximum Likelihood Classification

We now consider a richer problem, where the formation of the signal $\mathbf{x}$ under each hypothesis depends on specific parameters; this results in a combined detection and estimation problem. Specifically, for each class $\mathcal{C}_i$, an element $\mathbf{x} \in \mathcal{C}_i$ can be parameterized by a unique K-dimensional parameter vector $\boldsymbol{\theta}_i \in \Theta_i$ that controls the generation of the signal, i.e., $\mathbf{x} = f_i(\boldsymbol{\theta}_i)$ for some f_i. Example parameters for image classification scenarios include the pose of the object in the scene, translation, rotation, scale, etc.

In the case of two classes, where the optimal classifier is the likelihood ratio test, we can accommodate these unknown parameters through the use of the *generalized* likelihood ratio test. We will refer to the multi-class extension of this technique as the *generalized maximum likelihood classifier* (GMLC). To derive the GMLC, we again consider the case where noisy measurements of $\mathbf{x}$ are taken, $\mathbf{y} = \mathbf{x} + \boldsymbol{\omega}$, giving us a distribution $p(\mathbf{y}|\boldsymbol{\theta}_i, \mathcal{H}_i)$ for the measured signal $\mathbf{y}$ under hypothesis $\mathcal{H}_i$ *and* parameters $\boldsymbol{\theta}_i$. The GMLC is

$$\mathcal{C}(\mathbf{y}) = \underset{i=1,\ldots,P}{\arg\max} \ p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_i, \mathcal{H}_i), \tag{3}$$

where

$$\widehat{\boldsymbol{\theta}}_i = \underset{\theta \in \Theta_i}{\arg\max} \ p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}_i) \tag{4}$$

denotes the maximum likelihood estimate (MLE) of the parameters $\boldsymbol{\theta}_i$ under hypothesis $\mathcal{H}_i$. Under the same AWGN model for $\boldsymbol{\omega}$ as above, we have

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}_i) = \frac{1}{(2\pi\sigma)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma}\|\mathbf{y}-f_i(\theta)\|_2^2}, \tag{5}$$

and so the MLE (4) can be posed as

$$\widehat{\boldsymbol{\theta}}_i = \underset{\theta \in \Theta_i}{\arg\min} \ \|\mathbf{y} - f_i(\boldsymbol{\theta})\|_2^2. \tag{6}$$

3.3. GMLC and Smooth Manifolds

We can interpret the above situation geometrically. Consider the generation of the signal $\mathbf{x}$ under hypothesis $\mathcal{H}_i$, and recall that $\mathbf{x} = f_i(\boldsymbol{\theta}_i)$ for some K-dimensional parameter $\boldsymbol{\theta}_i \in \Theta_i$ and some function $f_i : \Theta_i \to \mathbb{R}^N$. If the mapping f_i is well-behaved, then the collection of signals $\mathcal{M}_i = \{f_i(\boldsymbol{\theta}_i) : \boldsymbol{\theta}_i \in \Theta_i\}$ forms a K-dimensional *manifold*[3] embedded in the ambient signal space $\mathbb{R}^N$. Under this interpretation, the ML estimation (6) of the parameter $\boldsymbol{\theta}_i$ under the AWGN model corresponds to finding the closest point on the manifold $\mathcal{M}_i$ to the observed signal $\mathbf{y}$. Subsequently, the classifier in (3) can be interpreted as a "nearest-manifold" search from $\mathbf{y}$ to each of the $\mathcal{M}_i$.

In settings where $\mathbf{x}$ is an image, the manifold $\mathcal{M}_i$ will have a degree of smoothness determined roughly by the spatial smoothness of the image $\mathbf{x}$.[9] If the manifold is differentiable, then ML estimation (6) can be performed using standard optimization techniques such as Newton's method. However, the manifolds generated by images having sharp edges are not differentiable, in general. In these cases, a coarse-to-fine differential estimation procedure[9] can be used for ML estimation that involves applying Newton's method to a multiscale sequence of smoothed versions of the image.

3.4. The Matched Filter

We conclude the discussion of traditional classification algorithms by recalling an important special case of the GMLC—the matched filter. While it is not typically presented in this manner, the matched filter is the optimal classifier when $f_i(t; \boldsymbol{\theta}_i) = \mathbf{s}_i(t - \boldsymbol{\theta}_i)$, i.e., our observed signal is one of a known set of possible signals $\mathbf{s}_i$ shifted by an unknown amount $\boldsymbol{\theta}_i$. The matched filter simply computes $\int \mathbf{x}(t)\mathbf{s}_i(t - \boldsymbol{\theta}_i)dt$ for all possible $\boldsymbol{\theta}_i$ and selects the class with the highest peak correlation value, which is equivalent to the GMLC when the signals $\mathbf{s}_i$ have equal norm for all i and additive white Gaussian noise is assumed. This can be efficiently implemented through convolution. Due to this simplicity, matched filtering techniques are frequently applied to classification problems even when the assumptions do not necessarily hold. For instance, in some image classification settings, the classifier must be accurate but also highly efficient since the classifier may function as part of a more complex, real-time system. In these cases, despite the wide variety of more sophisticated classification algorithms, it is common to use simple matched filters for image classification, where the signals $\mathbf{s}_i$ are constructed from a set of training images and are designed so that $\langle \mathbf{x}, \mathbf{s}_i \rangle$ is large when $\mathbf{x} \in \mathcal{C}_i$ and small otherwise.[10]

4. COMPRESSIVE CLASSIFICATION

In this section, we formulate a classification algorithm that uses compressive measurements to exploit the low-dimensional signal manifold structure that is present in many classification applications. In some cases we know the manifold structure explicitly, while in others we learn the manifold structure from training data, which serves as a sampling of points from each of the manifolds. This structure allows us to design efficient classification systems by reducing the dimension of the data required to perform the classification.

4.1. Maximum Likelihood Classification

We now consider the same classification problem as in Section 3.1, where each class corresponds to the presence of a known signal $\mathbf{s}_i \in \mathbb{R}^N$ in noise, but instead of observing $\mathbf{x} + \boldsymbol{\omega}$ we observe

$\mathbf{y} = \Phi(\mathbf{x} + \boldsymbol{\omega})$ where $\Phi \in \mathbb{R}^{M \times N}$, $M \leq N$. In this case, the MLC is essentially unchanged. In fact, when Φ is an orthoprojector and $\boldsymbol{\omega}$ is AWGN, the likelihood for each hypothesis $\mathcal{H}_i$ simply becomes

$$p(\mathbf{y}|\mathcal{H}_i) = \frac{1}{(2\pi\sigma)^{\frac{M}{2}}} e^{-\frac{1}{2\sigma}\|\mathbf{y} - \Phi\mathbf{s}_i\|_2^2}. \tag{7}$$

Thus, the MLC still reduces to nearest-neighbor classification among the available hypotheses, except in this case we compute $\|\mathbf{y} - \Phi\mathbf{s}_i\|_2$ (where $\mathbf{y} \in \mathbb{R}^M$) for each class rather than $\|\mathbf{y} - \mathbf{s}_i\|_2$ (where $\mathbf{y} \in \mathbb{R}^N$).

Two remarks are in order. First, the performance of the compressive MLC does not depend on any structure of the $\mathbf{s}_i$ such as sparsity or compressibility. Second, dimensionality reduction through a random orthoprojector will, with high probability, reduce the distance between two arbitrary points by a factor of approximately $\sqrt{M/N}$, while not affecting the variance of additive white Gaussian noise. Thus, the SNR of the projected signal is reduced compared to that in the ambient space by a factor of $\log(\sqrt{M/N})$. Hence, using the compressive MLC with $M \ll N$ may lead to an increase in the number of classification errors at high noise levels, and so M must be chosen with care.[6]

4.2. Generalized Maximum Likelihood Classification

We are now ready to describe our GMLC-based framework for compressive classification. Assume the parameterized, multiple hypothesis setting of Section 3.2, and suppose that we measure $\mathbf{y} = \Phi(\mathbf{x} + \boldsymbol{\omega})$. Again assuming that Φ is an orthoprojector, the likelihood for each hypothesis $\mathcal{H}_i$ simply becomes

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}_i) = \frac{1}{(2\pi\sigma)^{\frac{M}{2}}} e^{-\frac{1}{2\sigma}\|\mathbf{y} - \Phi f_i(\theta)\|_2^2}, \tag{8}$$

which in turn reduces the MLE to

$$\widehat{\boldsymbol{\theta}}_i = \arg\min_{\theta \in \Theta_i} \|\mathbf{y} - \Phi f_i(\boldsymbol{\theta})\|_2^2. \tag{9}$$

Thus, our classifier again reduces to a "nearest-manifold" classifier; the only significant difference is that the P classes now correspond to the manifolds $\Phi\mathcal{M}_i \subset \mathbb{R}^M$, $i = 1, \ldots, P$.[*]

As above, the performance of the GMLC does not depend on any structure of the signals $\mathbf{s}_i$. Rather, its performance depends on the stability of the dimensionality reduction of the manifold: if the distances between the projected points of the manifold and the projected signal are not preserved, then the estimator performance will suffer. This issue becomes critical during the nearest-neighbor classification step.

4.3. Stable Embedding of Multiple Smooth Manifolds

Let us now more closely examine the random projection of one or more manifolds from a high-dimensional ambient space $\mathbb{R}^N$ to a lower-dimensional subspace $\mathbb{R}^M$. We have recently shown[3] that this process actually preserves the essential structure of a *smooth* manifold, provided that a sufficient number M of random projections are taken (see Theorem 1 below). Just as the CS theory demands a projection dimension M proportional to the sparsity K, the requisite M to ensure a satisfactory embedding of a manifold depends on properties of that manifold. The primary dependence is

---

[*]Linear projection by Φ of a manifold $\mathcal{M} \in \mathbb{R}^N$ yields another manifold in $\Phi\mathcal{M} \in \mathbb{R}^M$.

on the dimension K of the manifold, but additional factors such as the volume and curvature of the manifold play a minor role. After stating the theorem, we continue our discussion of these properties.

THEOREM 1. [3] *Let $\mathcal{M}$ be a compact K-dimensional submanifold of $\mathbb{R}^N$ having condition number $1/\tau$ and volume V. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with*

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1})\log(1/\rho)}{\epsilon^2}\right). \tag{10}$$

If $M \leq N$, then with probability at least $1 - \rho$ the following statement holds: For every pair of points $x_1, x_2 \in \mathcal{M}$,

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}}.$$

The condition number $1/\tau$ in Theorem 1 is a measure of both the local curvature of the manifold and its global self-avoidance. Essentially, τ defines the maximum radius of a sphere that, when placed tangent to the manifold at any point, intersects the manifold only at that point; smaller τ leads to a larger condition number, indicating a less regular manifold.[3,11]

Theorem 1 states that, for sufficiently large M, all pairwise distances between points on a manifold are well-preserved under projection to a random M-dimensional subspace. (It follows also that geodesic distances are well-preserved in addition to the manifold's dimension, volume, topology, etc.[3]) By appropriate considerations of the properties $1/\tau$ and V this theorem can be extended to account for the simultaneous projection of multiple manifolds.

COROLLARY 1. *Let $\{\mathcal{M}_i\}_{i=1}^P$ be compact K-dimensional submanifolds of $\mathbb{R}^N$ having condition numbers $1/\tau_i$ and volumes V_i, respectively. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$ and let*

$$V = \sum_i V_i \quad \text{and} \quad \tau = \min\left(\min_i \tau_i, \ \min_{i \neq j} \ \text{dist}(\mathcal{M}_i, \mathcal{M}_j)\right).$$
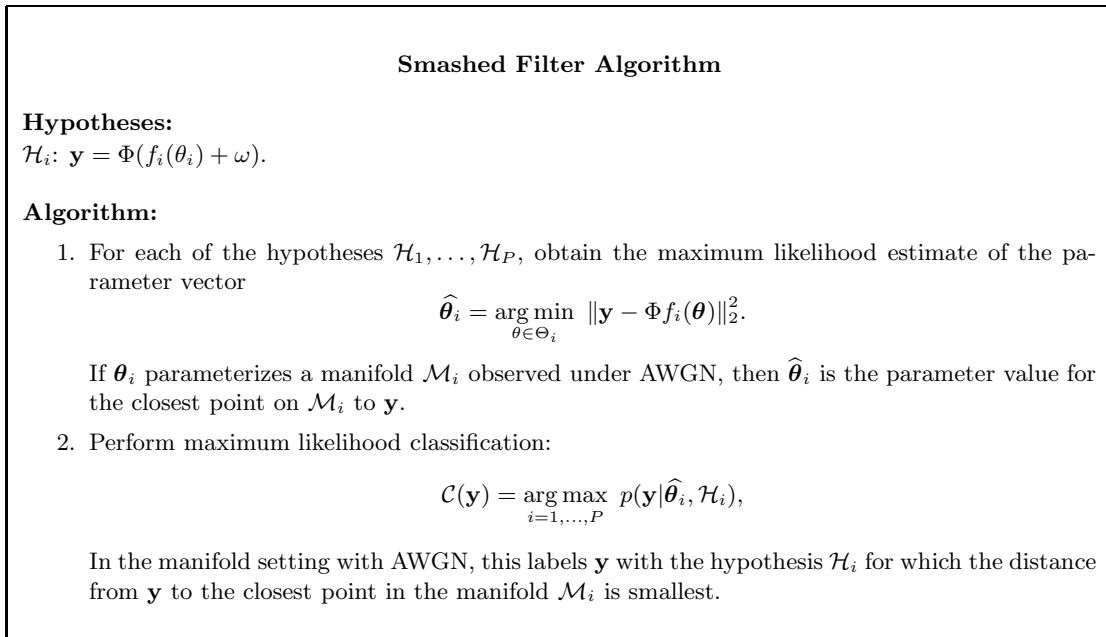
Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1})\log(1/\rho)}{\epsilon^2}\right). \tag{11}$$

If $M \leq N$, then with probability at least $1 - \rho$ the following statement holds: For every pair of points $x_1, x_2 \in \cup_i \mathcal{M}_i$,

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}}.$$

Corollary 1 ensures not only that distances between pairs of points on each manifold are well-preserved, but also that the distances between the P manifolds themselves are all well-preserved. The cost in terms of measurements is extremely modest; assuming similarly conditioned manifolds, the difference between (11) and (10) is approximately $O(K \log(P))$ additional measurements. In a classification setting with a large number of possible classes, this *sublinear* growth in the required number of measurements is particularly attractive.

Smashed Filter Algorithm

Hypotheses:
$\mathcal{H}_i$: $\mathbf{y} = \Phi(f_i(\theta_i) + \omega)$.

Algorithm:

1. For each of the hypotheses $\mathcal{H}_1, \ldots, \mathcal{H}_P$, obtain the maximum likelihood estimate of the parameter vector
$$\widehat{\boldsymbol{\theta}}_i = \arg\min_{\theta \in \Theta_i} \ \|\mathbf{y} - \Phi f_i(\boldsymbol{\theta})\|_2^2.$$
If $\boldsymbol{\theta}_i$ parameterizes a manifold $\mathcal{M}_i$ observed under AWGN, then $\widehat{\boldsymbol{\theta}}_i$ is the parameter value for the closest point on $\mathcal{M}_i$ to $\mathbf{y}$.

2. Perform maximum likelihood classification:
$$\mathcal{C}(\mathbf{y}) = \arg\max_{i=1,\ldots,P} \ p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_i, \mathcal{H}_i),$$
In the manifold setting with AWGN, this labels $\mathbf{y}$ with the hypothesis $\mathcal{H}_i$ for which the distance from $\mathbf{y}$ to the closest point in the manifold $\mathcal{M}_i$ is smallest.


Figure 2. Smashed filter algorithm.

Note that here the structure of the signals–namely that they inhabit low-dimensional manifolds in the high-dimensional ambient space–is critical to the performance of the GMLC. The individual signals, however, still do not need to have any sparsity or compressiblility properties; rather it is the manifolds that need to be compressible.

4.4. The Smashed Filter

In order to implement the GMLC described above, we first need to obtain estimates of the parameter vectors $\widehat{\boldsymbol{\theta}}_i$ from the noisy compressive measurements $\mathbf{y}$ under each of the hypotheses, which are used to classify the signal using the GMLC as described in Figure 2. We consider here some simple approaches to obtaining these estimates. Recall from Section 3.4 that the matched filter is an important special case of implementing the ML estimation step of the GMLC. To emphasize its similarity to the traditional matched filter while also stressing its *compressive* nature, we dub our approach the *smashed filter*.

First, suppose that we know the explicit structure of the manifold for each class, i.e., that we know the functions f_i. For example, each manifold might be parameterized by all possible shifts of a known 1-D signal or all possible positions of a known object in a 2-D image. In this setting we can explicitly search over the manifold using various approaches. In some cases we can use the same approach as the matched filter and simply continuously vary the manifold parameter(s), applying Φ to the resulting signals, and looking for the minimum distance between the output and the observed data. In the case where the original manifolds are differentiable, Corollary 1 ensures that the projected manifolds are also differentiable, and thus we can also use optimization techniques such as Newton's method to perform this optimization. In cases where the original manifolds are not differentiable, it is possible to construct measurement matrices that simultaneously smooth the

| (a) Tank | (b) School Bus | (c) Truck |
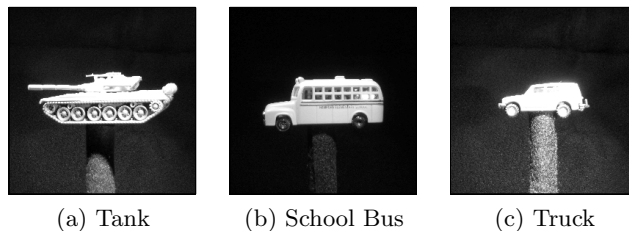
Figure 3. *Models used for classification experiments.*

data (at various resolutions) and apply a random projection.[12] This is beyond the scope of this paper but remains an interesting topic for further research.

We can also discretize an explicit search over the manifold using a grid-search procedure that, while not exhaustive, will yield a close approximation to the MLE. For example, consider the case where each class consists of a known 1-D signal $s_i(t - \boldsymbol{\theta}_i)$ with an unknown shift $\boldsymbol{\theta}_i$. In this case we can simply try a grid of possible values of $\boldsymbol{\theta}_i$ under each hypothesis assumption. This serves as a dense sampling of the manifold. Our ML estimate for $\boldsymbol{\theta}_i$ for each class is simply the value that minimizes $\|y - \Phi\widehat{\mathbf{s}}_i\|_2$, where $\widehat{\mathbf{s}}_i$ denotes a sampled version of $s_i(t - \widehat{\boldsymbol{\theta}}_i)$. The classification is then based on a nearest neighbor test among the points on each manifold selected as the MLE for that class. A similar approach could be taken in any case where we know the functions f_i.

In the case where we do not explicitly know the functions f_i for each class, but rather have training data from each class, the problem simplifies even further. We can simply assume that the training data points are drawn at random from the class-appropriate manifolds. Since our classifier ultimately depends on $f_i(\widehat{\boldsymbol{\theta}}_i)$, we do not need to know the mapping f_i. We can simply estimate $f_i(\widehat{\boldsymbol{\theta}}_i)$ as the nearest-neighbor from each class in the training set. We then proceed exactly as above.

4.5. Advantages of Compressive Classification

In addition to the computational and storage savings afforded by compressive classification, our proposed method shares many advantages previously shown for 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. In our context this means that we do not need to know what the classes are or what the classification algorithm will be prior to acquiring the measurements. Additionally, compressive measurements are *progressive* in the sense that larger numbers of projections translate into higher classification rates due to increased noise tolerance and *democratic* in that each measurement can be given equal priority because classification rates depend only on how many measurements are received, not on the particular subset received.

5. EXPERIMENTAL RESULTS

We now present results from a number of experiments that evaluate the smashed filter in a image target classification setting. We consider three classes, each for a different vehicle model: a tank, a school bus, and a truck (see Figure 3). All images are of size 128×128 pixels, hence $N = 16384$, and all measurement matrices are binary orthoprojectors obtained from a random number generator.

The first experiment is synthetic and concerns unknown shifts of a known image. In this case, $K = 2$ and we know the explicit structure of the three manifolds: each can be constructed by

translating a reference image in the 2-D image plane. The shifted versions of each image, as well as their corresponding compressive measurements, were obtained synthetically using software. The first step in implementing the smashed filter is thus to find the ML estimate of the shift for each class. This can be accomplished by simply calculating the distance between the observed **y** and projections of all possible shifts of the image. For simplicity, we assume that the set of possible shifts is limited by a maximum shift of 16 pixels in any direction. After the shift estimate is obtained for each class, the GMLC selects the class whose estimate is closest to the observed image.

We performed classification experiments for different numbers of compressive measurements, varying from $M = 2$ to 60, and for different levels of additive Gaussian noise, with standard deviations $\sigma = 0.001, 0.005, 0.01$ and 0.02. For each setting, we executed 10000 iterations of the experiment, where we selected a testing point at random with a different noise realization at each iteration. The plots of the average and minimum distances between manifolds in Figure 4(a) show that the distance is proportional roughly to the square root of the number of measurements M. In noisy measurement settings, a noise level similar to the average distance affects estimation and classification performance.

The classification rates in Figure 4(b) show a clear dependence between the number of measurements and the classification rate. Furthermore, performance for a given M degrades as the noise level increases; as expected the classifier becomes unreliable when the noise level becomes comparable to the minimum distance between the projected manifolds. In Figure 4(c), we see a similar relationship between the noise level and the error in the parameter estimate. These results verify that increasing the number of measurements improves the quality of the estimates; and the performance of the classifier is clearly dependent on the performance of the parameter estimator for the appropriate class.

The second experiment uses real data from the single-pixel camera described in Section 2 and concerns unknown rotations of the three targets in the z-axis in $\mathbb{R}^3$. In this case $K = 1$, and we assume that we do not know the explicit structure of the three manifolds; hence we use training data to provide an estimate of the manifold structure. We acquired a training set of compressive measurements for each vehicle for rotation angles that are multiples of $10°$ ($10°, 20°, \ldots, 360°$). We first estimated the most likely rotation angle for each class by computing the nearest neighbor from each class and then performed nearest neighbor classification.

We evaluate the performance of the smashed filter classifier using leave-one-out testing. The measurements for each rotation/class combination were classified using a smashed filter trained on all other available data points. We performed classification experiments for different numbers of compressive measurements, varying from $M = 2$ to 60. Table 1 provides confusion matrices for $M = 2, 4,$ and 6. The confusion matrices summarize the distributions for elements belonging to a given class (one per row: tank, school bus, or truck) being assigned a given class label (one per column: tank, school bus, or truck). The diagonal elements show the probabilities of correct classification for each of the classes. The matrices show that performance improves as M increases. Specifically, for $M \geq 6$, the classification rate remains at 100%. Figure 4(d) plots the average rotation estimate error as a function of the number of measurements, with behavior similar to that of Figure 4(b).

6. DISCUSSION AND CONCLUSIONS

In this paper, we have demonstrated that, thanks to the pronounced structure present in many signal classes, small numbers of nonadaptive compressive measurements can suffice to capture the
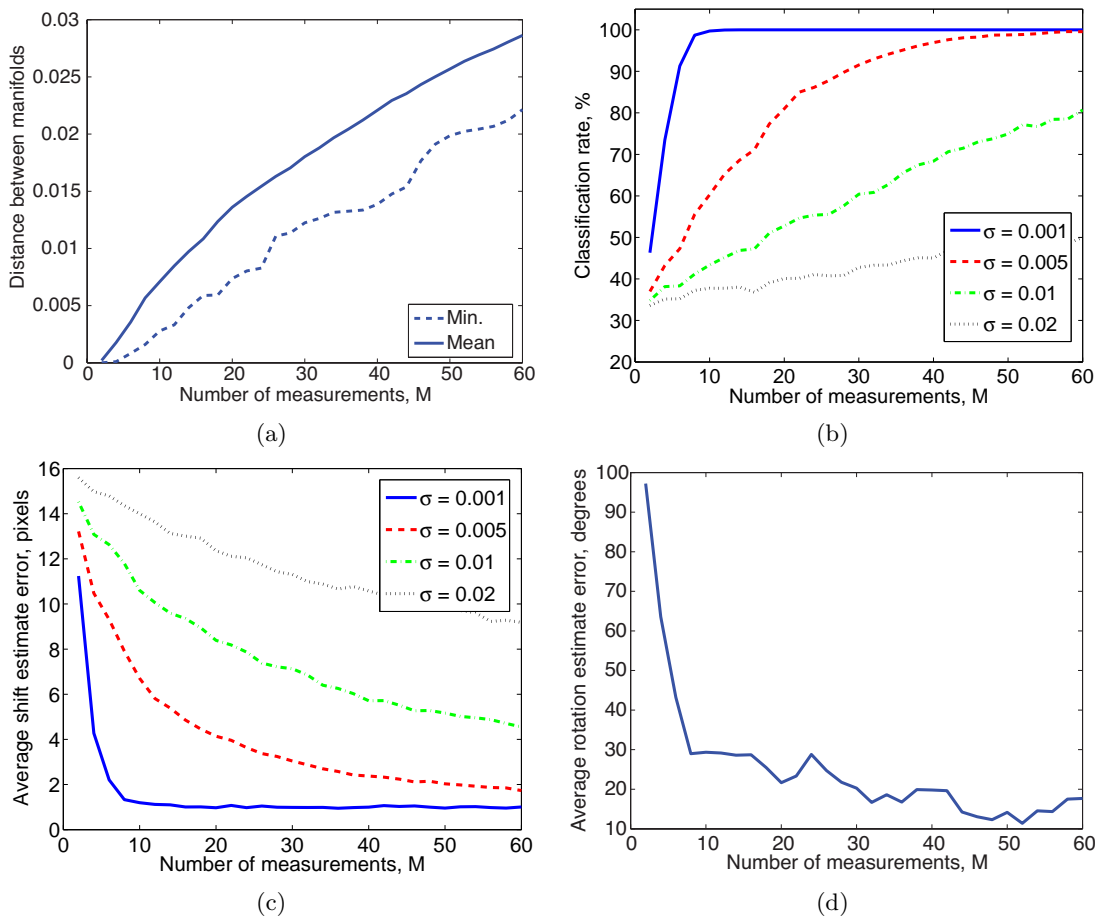
Figure 4. *Results for image classification experiments. (a) Minimum and mean distances between projected manifolds of shifted images; (b) classification rates and (c) average estimation error for varying number of measurements M and noise levels σ for the image shift experiments; (d) average estimate error for varying number of measurements M for the object rotation experiments. As M increases, the distances between the manifolds increase as well, thus increasing the noise tolerance and enabling more accurate estimation and classification. Thus, the classification and estimation performances improve as σ decreases and M increases in all cases.*

relevant information required for accurate classification. Our work rests on two key facts: (1) that simple parametric models impose a low-dimensional manifold structure on the signal classes within the high-dimensional image space, and (2) that the geometric structure of these manifolds is preserved under their projection to a random lower-dimensional subspace. The number of measurements required for a given classification performance level does not depend on the sparsity or compressibility of the images but only on the noise level and the structure of the manifolds, growing linearly in the dimensionality of the manifolds but only logarithmically in the number of pixels/samples and image classes. Our GMLC-based smashed filter is readily implementable with CS hardware such as the single-pixel compressive imaging camera and shares many of the attractive

Table 1. *Confusion matrices for rotation experiments for tank, school bus, and truck with varying number of measurements M.*

$M = 2$			$M = 4$			$M \geq 6$		
86.1%	0.0%	13.9%	94.4%	0.0%	5.6%	100%	0.0%	0.0%
0.0%	97.2%	2.8%	0.0%	100%	0.0%	0.0%	100%	0.0%
13.9%	0.0%	86.1%	16.7%	0.0%	83.3%	0.0%	0.0%	100%

features of CS in general, including simplicity, universality, robustness, democracy, and scalability, which should enable it to impact a variety of different applications.

In the future we hope to develop more sophisticated algorithms to exploit the manifold structure to more efficiently obtain the ML estimates required by the smashed filter. For example, rather than an exhaustive nearest-neighbor search, which could be computationally prohibitive for a large training set, a greedy approach might offer similar performance at significant computational savings; other approaches that exploit the smoothness of the manifolds could also be beneficial.

REFERENCES

1. E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Info. Theory* **52**, pp. 5406–5425, Dec. 2006.
2. D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory* **52**, pp. 1289–1306, Sept. 2006.
3. R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," 2006. Preprint.
4. D. Waagen, N. Shah, M. Ordaz, and M. Cassabaum, "Random subspaces and SAR classification efficacy," in *Proc. SPIE Algorithms for Synthetic Aperture Radar Imagery XII*, May 2005.
5. 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)*, **III**, pp. 305–308, (Toulouse, France), May 2006.
6. M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Detection and estimation with compressive measurements," Tech. Rep. TREE0610, Rice University ECE Department, 2006.
7. 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.
8. 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*, **6065**, pp. 43–52, (San Jose, CA), Jan. 2006.
9. M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *Proc. Wavelets XI at SPIE Optics and Photonics*, SPIE, (San Diego, CA), August 2005.
10. A. Mahalanobis, B. Kumar, and S. Sims, "Distance-classifier correlation filters for multiclass target recognition," *Applied Optics* **35**(17), pp. 3127–3133, 1996.
11. P. Niyogi, S. Smale, and S. Weinberger, "Finding the homology of submanifolds with confidence from random samples," *Discrete and Computational Geometry* , 2006. To appear.
12. M. B. Wakin, *The Geometry of Low-Dimensional Signal Manifolds.* PhD thesis, Rice University, Houston, TX, Aug. 2006.