Compressive Sensing Forensics

Xiaoyu Chu, *Student Member, IEEE*, Matthew C. Stamm, *Member, IEEE*, and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—Identifying a signal's origin and how it was acquired is an important forensic problem. While forensic techniques currently exist to determine a signal's acquisition history, these techniques do not account for the possibility that a signal could be compressively sensed. This is an important problem since compressive sensing techniques have seen increased popularity in recent years. In this paper, we propose a set of forensic techniques to identify signals acquired by compressive sensing. We do this by first identifying the fingerprints left in a signal by compressive sensing. We then propose two compressive sensing detection techniques that can operate on a broad class of signals. Since compressive sensing fingerprints can be confused with fingerprints left by traditional image compression techniques, we propose a forensic technique specifically designed to identify compressive sensing in digital images. Additionally, we propose a technique to forensically estimate the number of compressive measurements used to acquire a signal. Through a series of experiments, we demonstrate that each of our proposed techniques can perform reliably under realistic conditions. Simulation results show that both our zero ratio detector and distribution-based detector yield perfect detections for all reasonable conditions that compressive sensing is used in applications, and the specific two-step detector for images can at least achieve probability of detection of 90% for probability of false alarm less than 10%. Additionally, our estimator for the number of compressive measurements can well reflect the real number.**

I. Introduction

Since the initial development of digital multimedia forensics, researchers have sought to identify how different digital signals were captured and stored. Information about how a signal was acquired can be used to both identify the specific device used to capture the signal and to verify the signal's authenticity. Furthermore, knowledge of how a signal was captured can be used to help trace its processing history. As a result, determining how a signal was acquired has become an important forensic problem.

Typically, forensic algorithms determine how a signal was acquired by identifying imperceptible traces introduced into a digital signal during the acquisition process. These traces, which are known as fingerprints, arise due to properties of the sensor used to capture the signal or as a result of the signal processing operations used to form the digital signal. Existing forensic algorithms capable of identifying a signal's acquisition history are focused almost exclusively on images and videos [1]–[5]. While each of these specifically designed techniques performs strongly, it is necessary to develop forensic algorithms capable of identifying the acquisition history of a broader class of signals.

Recently, a new method of capturing signals known as *compressive sensing* has gained considerable attention. Compressive sensing is a signal processing technique capable of acquiring sparse signals at sampling rates below the Nyquist rate [6]. Rather than measuring the signal's value at a series of uniformly spaced points, each compressive measurement corresponds to a randomly weighted summation of the entire signal. The sparse signal can then be reconstructed using l_1 minimization from much fewer measurements than are needed by traditional uniform sampling [7]. Furthermore, many real signals that are not ideally sparse can be modeled as either sparse signals in the presence of noise or signals that are 'nearly sparse'. Compressive sensing can be used to acquire these signals with low amounts of reconstruction error [8].

Due to the effectiveness of compressive sensing's sub-Nyquist acquisition rate, researchers in various signal processing fields have applied compressive sensing techniques to many signal acquisition systems. These applicable fields include but not limited to magnetic resonance imaging [9], photoacoustic imaging [10], astronomical imaging [11], radar [12], electrocardiography [13], networked data [14], and speech and audio [15].

While acquisition schemes based on compressive sensing principles are widely studied in the realm of research, the impact of compressive sensing has led people to design and build real devices based on this technique. Single pixel or single sensor acquisition devices have been developed for capturing conventional images [16] and hyperspectral images [17]. In these applications, compressive sensing not only reduced the acquisition power but also solved the 'out of focus' problem encountered in traditional cameras [16]. Moreover, due to the power consumption of billions of A-to-D conversion in video acquisition, a custom CMOS chip was designed by adopting compressive sensing technology to slash energy consumption by a factor of 15 [18]. Devices that apply compressive sensing to other applicable signals have also been developed and built [19]. Researchers from Rice University have even started a company, called InView, to develop low cost shortwave infrared cameras using compressive sensing [20].

While an increasing number of technologies have begun to make use of compressive sensing, there are currently no existing forensic techniques capable of differentiating between signals captured using compressive sensing and those captured by traditional uniform sampling. This has important consequences for the forensics community.

As the number of devices that incorporate compressive sensing into their signal processing pipeline increases, detecting the use of compressive sensing will become an important part of forensically identifying a signal's origin. A motivating example can be seen in hyperspectral imaging, which is used in many critical applications such as surveillance drones and environmental monitoring. Compressive sensing has been recently used to capture and store hyperspectral images [21]. Detecting evidence of compressive sensing in a hyperspectral image can help forensic investigators identify the device. Furthermore, there may be scenarios where our government is presented with an image captured by another government's surveillance drone. In this scenario, we may want to analyze the image to 1) verify the validity of the image and 2) understand the capabilities of the other government's surveillance drone. Similarly, hyperspectral images of landscapes may potentially be used in court cases related to environmental contamination or mineral rights.

Additionally, the use of compressive sensing can affect the output of existing forensic algorithms. For example, compressive sensing may also be used to acquire, compress, and store certain types of images [21]. However, existing compression detection schemes in [22] and [23] may misidentify a com-

This work is supported in part by AFOSR grant FA95500910179.

Xiaoyu Chu and K. J. Ray Liu are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: {cxygrace, kjrliu}@umd.edu).

Matthew C. Stamm is with the Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA 19104 USA (e-mail: mstamm@coe.drexel.edu).

pressively sensed image as an image that has been captured by a standard digital camera, then subsequently compressed. Thus, it is necessary to design a specific forensic scheme for compressive sensing detection to solve such confusions. In summary, it is clear that the identification of compressively sensed signals is an important forensic problem.

In this paper, we propose a new forensic technique capable of identifying signals that have been acquired by compressive sensing. We begin by identifying the fingerprints that compressive sensing introduces into a signal. Because virtually no compressively sensed signal is truly sparse, we show that the reconstruction error introduced into compressively sensed signals has certain characteristics. We use these characteristics as compressive sensing's fingerprints and examine these fingerprints under three models commonly applied to compressively sensed signals: sparse signals in the presence of noise, nearly sparse signals, and nearly sparse signals in the presence of noise. We then propose a set of forensic techniques to identify compressively sensed signals that fit each of these models. Furthermore, we develop a forensic technique specifically designed to identify compressively sensed images and differentiate them from images that have undergone traditional lossy compression. Additionally, we propose a technique to forensically estimate the number of compressive measurements used to acquire a signal.

The remainder of this paper is organized as follows. In Section II, we provide a brief review of compressive sensing and present three different models of compressively sensed signals. In Section III, we identify and analyze the fingerprints left in a signal by compressive sensing. Using these fingerprints, we propose two different compressive sensing detection techniques in Section IV. To address specific challenges encountered when identifying compressive sensing in digital images, we present a two step compressive sensing detection technique that can discriminate between images that have been compressed using wavelet-based coders and images that have been compressively sensed in Section V. In Section VI, we propose an estimator for the number of compressive measurements used to acquire a signal. A series of experimental results are presented in Section VII that demonstrate the effectiveness of our proposed forensic techniques. Finally, in Section VIII we conclude this paper.

II. System Model

We begin this section by providing a brief overview of compressive sensing. We then discuss the three different models used for real world signals that are compressively sensed. Throughout this paper, we will use $\underline{s}$ and $\underline{x}$ to denote the original signal and the observed signal, respectively. Given the observed signal may be obtained by either traditional sensing or compressive sensing, it will correspondingly equal to the direct, maybe noisy, observation of the original signal, or the reconstructed one from compressive measurements.

A. Compressive Sensing Overview

Traditionally, a discretely indexed signal is formed from a continuously indexed signal through uniform sampling. During uniform sampling, observations of the continuously indexed signal are performed at uniformly spaced intervals over a fixed duration. As a result, each entry s_i in a discretely indexed signal $\underline{s} = (s_1, s_2, \ldots, s_n)^T$ corresponds to a single, direct measurement of the continuously indexed signal, and we directly observe these measurements in traditional sensing. Thus, if we use $\underline{x}$ to denote the observed signal in such case, then $\underline{x} = \underline{s}$.

The recent development of compressive sensing has allowed sparse signals, which have only a few nonzero entries, to be captured with far fewer observations than traditional sampling. During compressive sensing, each compressive measurement corresponds to a linear combination of the continuously indexed signal's values at all the locations that would be observed during uniform sampling. Defining the weighting vector for the i^{th} compressive measurement as $\underline{\varphi}_i$, then each compressive measurement y_i can be written as

$$y_i = \underline{\varphi}_i^T \underline{s}. \tag{1}$$

If $m (m \ll n)$ compressive measurements are collected, the transpose of the set of weighting vectors can be vertically concatenated to form the observation matrix Φ. As a result, the measurement vector $\underline{y} = (y_1, y_2, \ldots, y_m)^T$ containing each compressive measurement can be written as

$$\underline{y} = \Phi \underline{s}. \tag{2}$$

Typically, random matrices are used for observation matrices Φ in order to satisfy the restricted isometry property for later reconstruction [8]. In this paper, we use Gaussian distribution with zero mean and unit variance to generate matrix Φ.

After the compressive measurements are obtained, the discretely indexed signal $\underline{x}$, which we will observe from compressive sensing, is reconstructed from the compressive measurements. This is done by solving the following constrained l_1 minimization problem

$$\min_{\underline{\tilde{x}}} ||\underline{\tilde{x}}||_{l_1}, \quad s.t. \;\; \Phi \underline{\tilde{x}} = \underline{y}. \tag{3}$$

If $\underline{s}$ is sparse, then given enough compressive measurements, $O(k \log n)$, where k and n are the sparsity and length of $\underline{s}$ respectively, the signal can be perfectly reconstructed, i.e. $\underline{x} = \underline{s}$ [7].

Compressive sensing forensics, however, is a reverse engineering problem of compressive sensing, which starts from the reconstructed signal and tries to reveal how the signal was acquired. Forensic investigators only observe a reconstructed signal $\underline{x}$. Then, based on the fingerprints extracted from this signal, they identify whether the observed signal was traditionally sensed or compressively sensed and reconstructed. Furthermore, forensic investigators can also estimate the number of compressive measurements m solely based on the reconstructed signal.

B. Signal Model

In theory, if a truly sparse signal is compressively sensed, it can be perfectly reconstructed [7]. In practice, however, this is rarely the case. Often, the compressive measurements of a truly sparse signal will be corrupted by noise. This can occur due to sensing in a noisy environment or due to noise within the sensors themselves. Furthermore, it is often the case that signals of interest are not truly sparse, but rather nearly sparse or 'compressible'. While non-sparse but compressible signals cannot be perfectly reconstructed, a bound can be placed on the reconstruction error [8]. If enough compressive measurements are captured, the reconstruction error can be made sufficiently small.

Here, we discuss several commonly used models applied to signals that are compressively sensed in real world scenarios. In subsequent sections, we will exploit the effects of these nonideal conditions to identify the use of compressive sensing.

Sparse Signals in the Presence of Noise

There are many scenarios in which a true signal has only a few nonzero coefficients (i.e., nonzero entries s_i in $\underline{s}$), but the signal is corrupted by noise during sensing. These signals can be modeled as sparse signals in the presence of noise. For

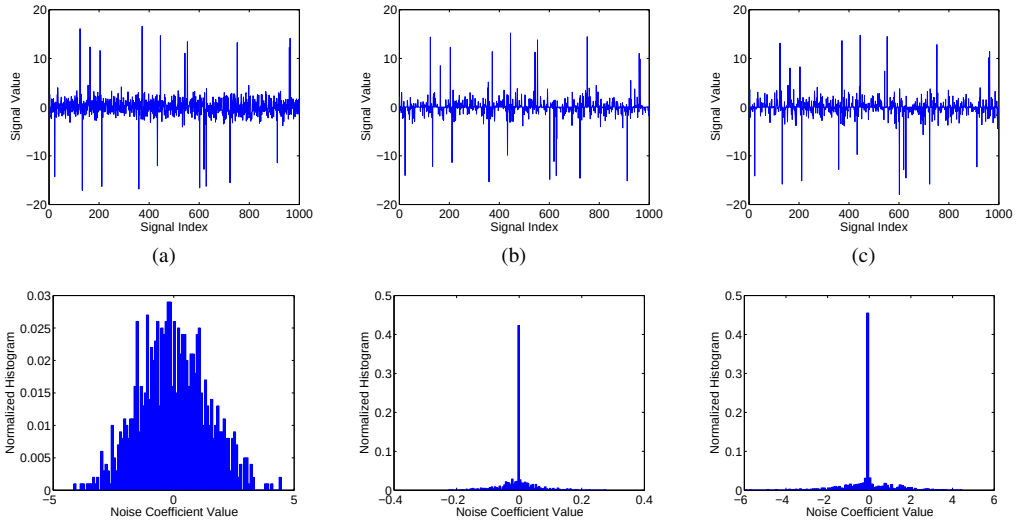


Fig. 1. Fingerprints of compressive sensing for sparse signals in the presence of measurement noise or environment noise. The upper row shows the observed signals from (a) traditional sensing, (b) compressive sensing corrupted with measurement noise and (c) compressive sensing corrupted with environment noise. The bottom row shows the corresponding noise histograms of the observed signals above.

example, in radar signal analysis the time-frequency plane is discretized into a grid where the number of grid cells is much larger than the total number of targets. The radar coefficients under this time-frequency shift operator basis are modeled as sparse signals in the presence of noise [24].

Under this model, let $\underline{s}$ represent a sparse signal to be sensed. If $\underline{s}$ is sensed using traditional uniform sampling, the observed signal $\underline{x}$ is given by

$$\underline{x} = \underline{s} + \underline{\eta}. \tag{4}$$

where $\underline{\eta}$ is a vector containing i.i.d. noise. Regardless of whether the noise originates in the sensor or is due to an environmental source, a unique noise measurement occurs at each signal observation x_i.

If $\underline{s}$ is compressively sensed, however, noise can be introduced into the compressive measurements. Under some scenarios, additive noise directly corrupts each compressive measurement [24]. This is equivalent to sensing using a noisy sensor. We refer to this type of noise as measurement noise, and model the compressive measurements as

$$\underline{y} = \Phi\,\underline{s} + \underline{\eta}^m, \tag{5}$$

where $\underline{\eta}^m$ is i.i.d. noise. In other scenarios, the sparse signal directly mixes with some noise process while it is being sensed [15]. We refer to this type of noise as environment noise. We model compressive measurements in the presence of i.i.d. environment noise $\underline{\eta}^e$ as

$$\underline{y} = \Phi(\underline{s} + \underline{\eta}^e). \tag{6}$$

If the compressive measurements are corrupted by either measurement or environment noise, the sparse signal is no longer reconstructed using (3). Instead, the reconstructed signal $\underline{x}$ is obtained by solving

$$\min_{\tilde{\underline{x}}} ||\tilde{\underline{x}}||_{l_1}, \quad s.t. \quad ||\underline{y} - \Phi\tilde{\underline{x}}||_{l_2}^2 \leq \epsilon \tag{7}$$

where ϵ is a parameter that depends on the noise power [25]. We note that in this equation, the constraint present in (3) is replaced with the inequality $||\underline{y} - \Phi\tilde{\underline{x}}||_{l_2}^2 \leq \epsilon$.

Nearly Sparse Signals

While many important types of signals are not truly sparse, they satisfy certain conditions allowing them to be well approximated by sparse signals. These signals are known as nearly sparse or compressible signals. The discrete wavelet transform coefficients of a digital image corresponding to a natural scene are a widely used example of a nearly sparse signal [26]. Gabor coefficients of certain classes of oscillatory signals can also be modeled as nearly sparse signals [27]. Though nearly sparse signals cannot be perfectly reconstructed if they are compressively sensed, they can be reconstructed with little error if enough compressive measurements are obtained.

To formally define nearly sparse signals, we first sort the entries of the signal $\underline{s}$ in descending order $s_{(1)}, s_{(2)}, \ldots, s_{(n)}$, such that $|s_{(1)}| \geq |s_{(2)}| \geq \ldots \geq |s_{(n)}|$. The signal $\underline{s}$ is compressible if and only if its sorted coefficients fall inside a weak l_p ball of radius R for some $0 < p < \infty$ [8], i.e.

$$|s_{(i)}| \leq R \cdot i^{-1/p}, \quad i = 1, 2, \ldots, n. \tag{8}$$

We model nearly sparse signals as compressible signals whose entries are i.i.d. random variables. Signals drawn from many commonly occurring distributions such as the Laplace and Gaussian distributions are compressible [8].

Nearly Sparse Signals in the Presence of Noise

In some real world scenarios, a nearly sparse signal may be compressively sensed in a noisy environment. As a result, we adopt nearly sparse signals in the presence of noise as a third signal model. These signals can be viewed as a combination of the previous two models. Provided that the noise power is sufficiently small, nearly sparse signals will remain compressible when corrupted by noise. As a result, we will see that detecting compressive sensing in signals that fit this models is similar to detecting compressive sensing in nearly sparse signals.

III. Compressive Sensing Fingerprints

To identify the fingerprints left by compressive sensing, we first examine sparse signals in the presence of noise, then examine nearly sparse signals.

Consider a signal $\underline{x}$ formed by sensing a sparse signal $\underline{s}$ in the presence of noise. Assuming that the locations of the nonzero components of $\underline{s}$ are known, the entries of $\underline{x}$ that do not correspond to nonzero values can be gathered together to form the vector $\underline{x}^n$. If $\underline{x}$ was acquired using traditional uniform sampling, each entry in $\underline{x}^n$ will directly correspond to a single noise observation. As a result, the normalized histogram of $\underline{x}^n$ approximates the distribution of the noise source. This can be seen in Fig. 1(d).

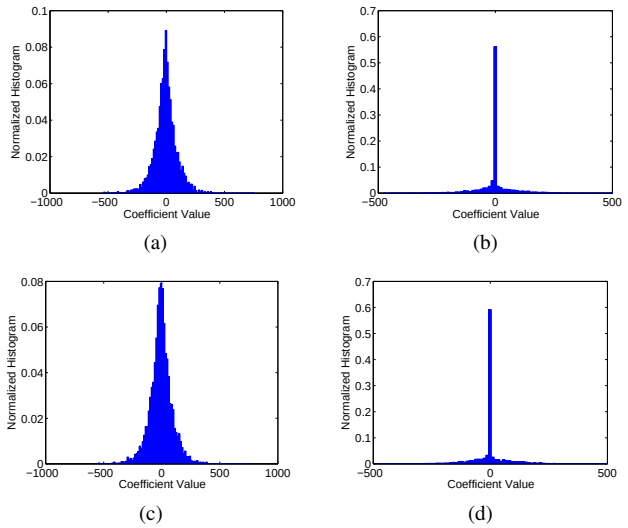


Fig. 2. Example showing the fingerprints of compressive sensing in a nearly sparse signal with and without the presence of noise. The top row shows the histograms observed from a nearly sparse signal after (a) traditional sensing and (b) compressive sensing. The bottom row shows the histograms observed from a nearly sparse signal in the presence of noise after (c) traditional sensing and (d) compressive sensing.

This is not the case, however, if $\underline{x}$ was acquired via compressive sensing. If measurement noise is encountered during sensing, the noise affects each compressive measurement. During reconstruction, no single value of $\underline{x}$ will correspond to a single noise observation. If environment noise is present during compressive sensing, both the sparse signal and the noise will be captured during the measurement process. Reconstructing the signal by solving (7), however, ensures that $\underline{x}$ will accurately reconstruct the $\underline{s}$ but not the noise. As a result, if $\underline{x}$ was captured using compressive sensing, the normalized histogram of $\underline{x}^n$ will not match the distribution of the noise source. In fact, because $\underline{x}$ was chosen to maximize the sparsity of the reconstructed signal, a significant number of entries in $\underline{x}^n$ will be zero or near zero. This will result in the presence of an impulsive peak at zero in the normalized histogram of $\underline{x}^n$ as can be seen in Fig.s 1(e) and (f). This peak is the fingerprints left by compressive sensing for sparse signals in the presence of noise.

A similar effect can be observed if $\underline{x}$ was formed by sensing a nearly sparse signal. As it is shown in Fig. 2(a), the normalized histogram of traditionally sensed signal $\underline{x}$ will closely match the distribution of the nearly sparse signal being sensed. However, the use of compressive sensing will greatly increase the histogram's kurtosis and result in a big concentration at zero as can be seen in Fig. 2(b). Furthermore, this result holds true for nearly sparse signals in the presence of noise, as can be seen in Fig. 2(c) and (d).

To show the effectiveness of compressive sensing fingerprints in real applications, we take a hyperspectral image, which is shown in Fig. 3(a), as an example. Hyperspectral images are composed of many sub-images in different spectrum bands, each of which can be obtained by compressive sensing [17]. Therefore, in this example, we take one sub-image out to examine. Comparing the traditionally sensed sub-image in Fig. 3(b) and the compressively sensed image in Fig. 3(c), we can hardly tell the difference. However, the histogram of transform domain coefficients from compressively sensed image, as it is shown in Fig. 3(d), has a much higher kurtosis at zero than that from the traditionally sensed image, which is shown in Fig. 3(e).

Furthermore, in order to show that such fingerprints also exist in real compressive sensing devices, we examine a single pixel

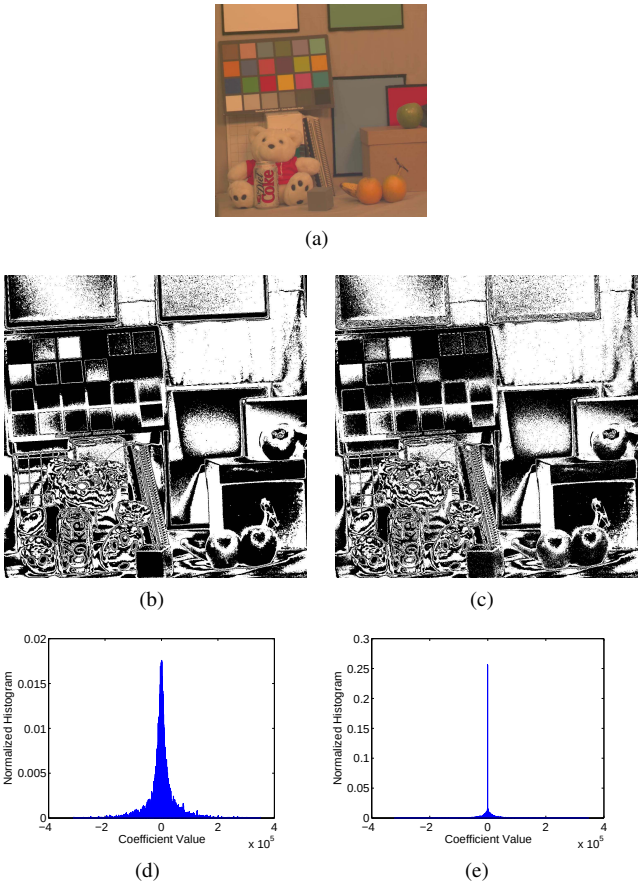
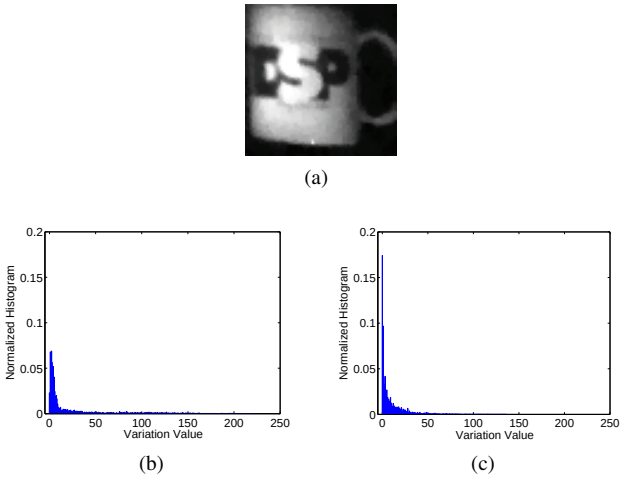


Fig. 3. (a) A hyperspectral image taken from [28] with dimension 1024×1024 pixel. (b) It's monochromatic image (obtained from raw data) corresponding to wavelength of 400nm. (c) The same monochromatic image obtained by compressive sensing and reconstructed from $1024^2 \times 50\%$ compressive measurements. (d) and (e) Histograms of DWT subband 3 coefficients from (d) the traditionally sensed image and (e) the compressively sensed image.

camera captured image and an image of the same scene but being captured by a traditional digital camera [29]. The single pixel camera in [29] obtains each compressive measurement by projecting the scene onto a randomized digital micromirror array and optically calculate the linear combination. We use the 'mug' image captured by a single pixel camera in [29], as it is shown in Fig. 4(a), to present the fingerprints of compressive sensing. Because the reconstruction step was performed by minimizing the total variation, the domain that compressive sensing fingerprints are present in is the pixel variations, i.e., gradient magnitudes. Fig.s 4(b) and 4(c) show the histograms of pixel variations for the traditionally sensed 'mug' image, and its compressively sensed version, respectively. We can see from Fig. 4(c) that a peak corresponding to a large concentration of components is present at the zero bin for the compressively sensed image. These fingerprints are absent from the traditionally captured image's histogram on the left.

We note that the compressive sensing fingerprints' existence is due to the sparse representation of the signal created upon reconstruction. Because all reconstruction algorithms enforce sparsity in one way or another, these fingerprints will be present in the sparsity domain regardless of the reconstruction algorithm.

Though we focus on the basis pursuit (BP) reconstruction algorithm in this paper, we note that there are several algorithms that can be used to reconstruct a compressively sensed signal such as orthogonal matching pursuit (OMP) [30], least absolute shrinkage and selection operator (LASSO) [31], and total variation (TV) [32]. We note that as long as a



(a) (b) (c)

Fig. 4. An example showing compressive sensing fingerprints in the (a) 'mug' image captured by a single pixel camera [29]. (b) The histogram of pixel variations (magnitude of the gradient) for the 'mug' image captured by a traditional digital camera. (c) The pixel variation histogram for the compressively sensed image of the same scene acquired using the single pixel camera.

reconstruction algorithm seeks a sparse representation of the compressive measurements, similar fingerprints will be present in the reconstructed signal.

IV. Compressive Sensing Detection

Now that we have identified the fingerprints left by compressive sensing, we are able to develop a set of forensic techniques to detect its use [33]. Detecting the use of compressive sensing is equivalent to differentiating between the following hypotheses

$$\begin{aligned} H_0: &\quad \underline{x} \text{ was obtained using traditional sampling,} \\ H_1: &\quad \underline{x} \text{ was obtained using compressive sensing,} \end{aligned} \tag{9}$$

where $\underline{x}$ is a discretely indexed signal of unknown origin. To do this, we first need to obtain some measure of the strength of any compressive sensing fingerprints present in $\underline{x}$. Measurement of these fingerprints' strength, however, depends on the appropriate signal model for $\underline{x}$ as well as the amount of side information known by the forensic investigator. To account for this, we propose two different compressive sensing detection techniques that are appropriate in different forensic scenarios.

A. Zero Ratio Detection Scheme

In many cases, a forensic investigator knows little more than the fact that the signal in question fits one of the three signal models outlined in Section III. If this is the case, the forensic investigator cannot leverage any side information such as the signal or noise distribution while measuring the strength of compressive sensing fingerprints. The investigator can, however, make use of the fact that if compressive sensing was performed, it was done under nonideal conditions.

Assume temporarily that $\underline{x}$ can be modeled as a sparse signal $\underline{s}$ sensed in the presence of noise. We assume that the noise has a continuous distribution and a nonzero variance, i.e. its distribution is not an impulse. From Section III, we know that under hypothesis H_0 each entry of $\underline{x}^n$ will correspond directly to a noise observation. As a result, the distribution of the entries in $\underline{x}^n$ will match the noise distribution. By contrast, under hypothesis H_1, an impulsive peak located at zero will occur in the distribution of the entries of $\underline{x}^n$. Because of this, we can state

$$\mathbb{P}(x_i^n = 0|H_0) \ll \mathbb{P}(x_i^n = 0|H_1). \tag{10}$$

Though a forensic investigator may not know the noise distribution, the investigator can use (10) to measure the strength of compressive sensing fingerprints by calculating the ratio of zero valued entries in $\underline{x}^n$ to its total length.

Since in practice many of the techniques used to solve (3) or (7) result in values of $\underline{x}^n$ close to but not exactly equal to zero, we measure the strength of the fingerprints as follows. Let $\Lambda_\varepsilon(\underline{x}^n)$ denote the number of elements in $\underline{x}^n$ which have an absolute value no greater than ε. We calculate the zero ratio fingerprints' strength using the equation

$$\xi_z(\underline{x}^n) = \frac{\Lambda_\varepsilon(\underline{x}^n)}{\ell(\underline{x}^n)}, \tag{11}$$

where $\ell(\underline{x}^n)$ is the length of the vector $\underline{x}^n$. When calculating Λ_ε, ε is chosen to be $\varepsilon = ||\,\underline{x}^n\,||_\infty/\alpha$, where α is a parameter that controls the range of values of $\underline{x}^n$ that are counted as zeros. Experimentally, we have observed that choosing $\alpha = 100$ yields desirable results. We then perform compressive sensing detection using the following decision rule

$$\delta_z = \begin{cases} H_0 & \text{if } \xi_z(\underline{x}^n) < \tau_z, \\ H_1 & \text{if } \xi_z(\underline{x}^n) \geq \tau_z. \end{cases} \tag{12}$$

where τ_z is a decision threshold.

In reality, the locations of the nonzero values of $\underline{s}$ may not be known to a forensic investigator, thus making it difficult to form $\underline{x}^n$ from $\underline{x}$. In this scenario, two approaches can be taken to perform compressive sensing detection. Since $\underline{s}$ will contain a small number of nonzero entries, entries in $\underline{x}$ corresponding to these entries in $\underline{s}$ will have values significantly larger in magnitude than the rest. In the first approach, if the entries of $\underline{x}$ are sorted in descending order, a substantial drop in the values of the entries of $\underline{x}$ will be observed when transitioning between nonzero entries of $\underline{s}$ and $\underline{x}^n$. Using this information, a threshold can be chosen to separate out $\underline{x}^n$ for use in detection. If a suitable threshold cannot be chosen to separate out $\underline{x}^n$, a second approach can be used. In this approach, $\underline{x}$ can be used instead of $\underline{x}^n$ in the detection algorithm. Since $\underline{s}$ will have few nonzero entries, the statistics of $\underline{x}^n$ will dominate and there will be little effect on the detection results.

Additionally, if $\underline{x}$ can be modeled as a nearly sparse signal or a nearly sparse signal in the presence of noise, the preceding detection technique can still be used, albeit with slight modification. From Section III, we know that for nearly sparse signals or nearly sparse signals in the presence of noise, the reconstruction step in compressive sensing will result in the presence of a large number of zero or near zero valued entries in $\underline{x}$. As a result, we can state

$$\mathbb{P}(x_i = 0|H_0) \ll \mathbb{P}(x_i = 0|H_1). \tag{13}$$

for nearly sparse signals and nearly sparse signals in the presence of noise. If we substitute $\underline{x}$ for $\underline{x}^n$ in equations (11), compressive sensing can be detected in nearly sparse signals using the decision rule δ_z presented in (12).

B. Distribution-based Detection Scheme

In some scenarios, the forensic investigator will have knowledge about the distribution $\mathcal{F}$ of the noise present during sensing, like the quantization noise [34], or about the distribution $\mathcal{G}$ of the coefficients in a nearly sparse signal. This knowledge can be used as side information to perform improved compressive sensing detection. To develop a detection scheme that makes use of this distribution information, let us examine the case of nearly sparse signals.

Let us assume that a forensic examiner knows that the coefficients of a nearly sparse signal are distributed according to some parametric distribution $\mathcal{G}(\theta)$, where the true value of the parameter θ is unknown. Additionally, assume that the forensic investigator knows an estimator $\hat{\theta}$ for the parameter

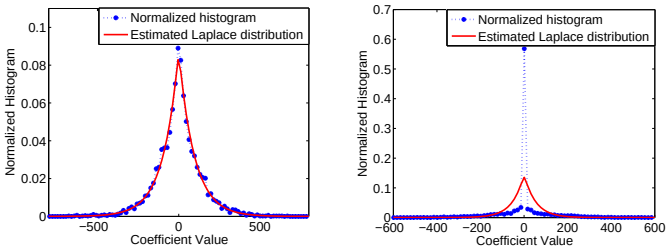


Fig. 5. Fitting the histogram of the observed signal to the estimated signal distribution. The Laplace distribution was used to generate each sample of the nearly sparse signal. The left figure shows the fitting result when this signal was obtained by traditional sensing, while the right one shows the result for a when the signal was compressively sensed.

θ on the basis of i.i.d. realizations of $\mathcal{G}(\theta)$. Under hypothesis H_0, each entry of $\underline{x}$ will be a direct observation of the nearly sparse signal, therefore the entries of $\underline{x}$ will be distributed according to $\mathcal{G}(\theta)$. If $\hat{\theta}$ is calculated using the entries of $\underline{x}$, an appropriately chosen measure of the distance between $\mathcal{G}(\hat{\theta})$ and the normalized histogram of $\underline{x}$ should be small. We know from Section III, however, that under hypothesis H_1 the entries of $\underline{x}$ will no longer be distributed according to $\mathcal{G}(\theta)$. This will cause $\hat{\theta}$ to be an inaccurate estimate of θ if it is calculated from $\underline{x}$ under hypothesis H_1. Now, given an appropriately chosen distance metric, the distance between $\mathcal{G}(\hat{\theta})$ and the normalized histogram of $\underline{x}$ will be large. This can be seen in Fig. 5. As a result, we can measure the strength of compressive sensing fingerprints in $\underline{x}$ by measuring the distance between the normalized histogram of $\underline{x}$ and $\mathcal{G}(\hat{\theta})$.

A problem arises when measuring the distance between these two quantities: $h_k(\underline{x})$ is an estimate of the probability that the value of x_i falls within the k^{th} histogram bin, while $\mathcal{G}(\hat{\theta}, t)$ is the probability that x_i takes the value t. As a result, these two quantities cannot be compared directly by any distance measurement. To resolve this disparity, we integrate $\mathcal{G}(\hat{\theta}, t)$ over each histogram bin to obtain $g(\hat{\theta})$ where

$$g_k(\hat{\theta}) = \int_{b(k-1/2)}^{b(k+1/2)} \mathcal{G}(\hat{\theta}, t)dt \tag{14}$$

and b is the width of each histogram bin.

Let $\xi_d(h_k, g_k)$ denote some distance measure between h_k and g_k, such as mean square distance (MSD) or Kullback-Leibler divergence (KL divergence), then, we perform compressive sensing detection using the following decision rule

$$\delta_d = \begin{cases} H_0 & \text{if } \xi_d(h_k, g_k) < \tau_d \\ H_1 & \text{if } \xi_d(h_k, g_k) \geq \tau_d. \end{cases} \tag{15}$$

where τ_d is a decision threshold. The choice of the distance measure $\xi_d(h_k, g_k)$ is made based on the performance of this compressive sensing detector in different applications. For example, when detecting compressively sensed images, using mean square error as the distance measure yield the best detection performance. We will discuss this case in the next section.

Besides the conventional distance measures, such as MSD and KL divergence, we also propose their modified versions as the candidates of $\xi_d(h_k, g_k)$. These modified distance measures take into account the particular manner in which compressive sensing changes the distribution of the entries in $\underline{x}$. Take the KL divergence measure as an example. Since compressive sensing dramatically increases the kurtosis of the distribution of the entries in $\underline{x}$, the most forensically significant differences between h and g should occur around $k = 0$. As a result, we modify the KL divergence to measure the strength of compressive sensing fingerprints as follows

$$\xi_d(h_k, g_k) = \sum_k w_k \ln \frac{h_k}{g_k}, \tag{16}$$

where w_k is a normalized set of weights used to emphasize differences in the forensically significant region around $k = 0$. Since we wish to weight the regions around $k = 0$ more heavily, we construct the weighting function using a Laplace distribution. Other distributions obeying power law decay may also be good candidates. Given that the weights are discrete, we integrate the Laplace distribution over each histogram bin to obtain the weighting function as follows,

$$w_k = \begin{cases} 1 - e^{-\nu b/2}\cosh(\nu k) & \text{if } k = 0, \\ e^{-\nu|k|}\sinh(\nu b/2) & \text{otherwise,} \end{cases} \tag{17}$$

where the parameter ν is chosen to be

$$\nu = \frac{\beta n}{\sum_{i=1}^{n} |x_i|}, \tag{18}$$

and where β is a user specified parameter that adjusts the size of the forensically significant region. Experimentally, we have found that $\beta = 100$ yields desirable results. Similar modifications can be applied on other conventional distance measures.

If the signal being examined can be modeled as a sparse signal in the presence of noise and the forensic investigator has a parametric model $\mathcal{F}(\theta)$ of the noise distribution, the detection technique presented above can be used, only with slight modifications. Since the noise distribution rather than the signal distribution is known, $\mathcal{F}$ should be substituted for $\mathcal{G}$ in (14). Additionally, $\hat{\theta}$ should be calculated using $\underline{x}^n$ and the histogram of $\underline{x}^n$ should be substituted for $h(\underline{x})$ in (16). If the signal is more appropriately modeled as a nearly sparse signal in the presence of noise, the distribution of $\underline{x}$ is given by the convolution of $\mathcal{G}$ and $\mathcal{F}$. If the noise distribution is unknown or if $\mathcal{G} * \mathcal{F}$ is difficult or intractable, the noise distribution can be ignored when performing compressive sensing detection as long as the noise power is sufficiently low.

We note that, although only the original signal's distribution is explicitly used in this distribution-based detection scheme, our model for compressively sensed signals has also been implicitly applied when designing the detector. Specifically, both detection schemes are designed based on the assumption that the distribution of a compressively sensed signal has much more kurtosis than that of a traditionally sensed signal. While this is enough for identifying compressively sensed signals from traditionally sensed signals, more explicit models for the distribution of compressively sensed signals can be proposed for particular applications where more complicated detection scenarios exist. We will discuss this in detail for images in the next section.

V. Detecting Compressive Sensing in Digital Images

While the compressive sensing detection techniques proposed in Section IV can be used on a wide variety of signals, in some scenarios it is desirable to create a compressive sensing detection technique specifically tailored to a particular class of signals. This is the case for digital images.

An image's compression history can reveal important information about how an image was captured and stored. It can also reveal important information about the device used to capture an image [3]. As a result, a variety of techniques have been developed to determine if an image was previously compressed. Fingerprints left by compressive sensing, however, can be mistaken for traditional image compression fingerprints by existing

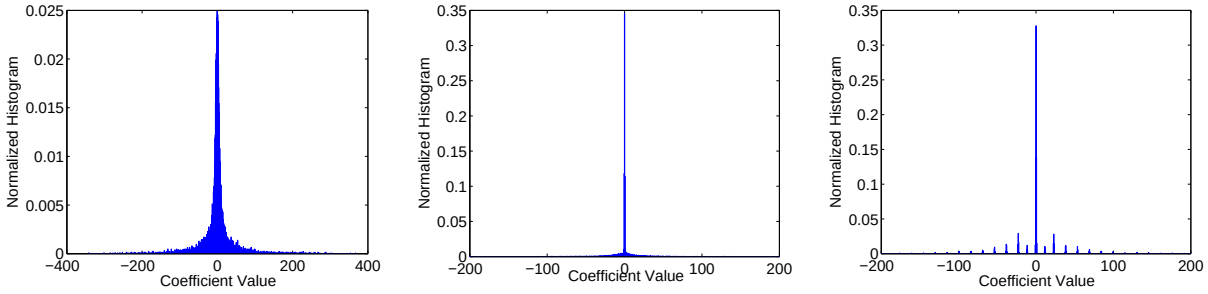


Fig. 6. Histograms of DWT coefficients taken from uncompressed Lena (left), the same image after JPEG 2000 compression (right), and the reconstructed compressively sensed Lena (center).

forensic techniques such as those proposed by Lin et al. [22] and Luo et al. [23]. As a result, when we are given a compressively sensed and reconstructed image, it may be easily misidentified as a traditionally sensed and compressed image. In this section, we propose a forensic technique specifically designed to both detect evidence of compressive sensing in digital images and to differentiate compressive sensing fingerprints from those left by traditional forms of image compression.

A. Compressive Sensing Fingerprints in Digital Images

Since the pixel values of an image do not form a sparse signal, digital images may not initially seem well suited for compressive sensing. It is well known, however, that within each subband, the set of discrete wavelet transform (DWT) coefficients of a natural image are sparse. As a result, compressive sensing reconstruction is often performed on images in the wavelet domain.

From our discussion of compressive sensing fingerprints in Section III, we would naturally expect an impulsive peak to occur at zero in the DWT coefficient distribution of a compressively sensed image. While this is true after the compressively sensed DWT coefficients are reconstructed, the inverse DWT of the image must be performed and the resulting pixel values must be projected back into the set $\{0, \ldots, 255\}$ of allowable pixel values. This will introduce a small but nontrivial amount of noise into the DWT coefficients when DWT is applied to the image again to extract the coefficients. As a result, the peak in the image's DWT coefficient distribution at zero will no longer correspond to an impulse. Though the peak will be slightly smoothed by this noise source, the DWT coefficient distribution of a compressively sensed image will still exhibit a large degree of kurtosis, as can be seen in Fig. 6. We use this characteristic feature of a compressively sensed image's DWT coefficient distribution as the fingerprints.

Wavelet-based image compression techniques such as JPEG 2000 and SPIHT also introduce fingerprints in an image's DWT coefficient distribution. During compression, these techniques use a bit-plane encoder to store the most significant digits of each DWT coefficient in a subband. This has the same effect as quantizing each DWT coefficient. As a result, the DWT coefficients in an image compressed using a wavelet-based technique will tightly cluster around certain values, forming a series of peaks in the DWT coefficient distribution that can be seen in the rightmost plot in Fig. 6. These peaks are the fingerprints of wavelet based image compression. Since the most prominent peak occurs at zero, compressive sensing fingerprints and wavelet-based compression fingerprints can easily be confused by existing detectors.

To demonstrate that compression history detection techniques can mistake compressive sensing fingerprints for JPEG 2000 compression fingerprints, we performed an experiment using

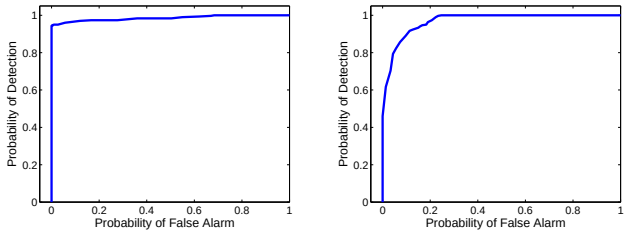


Fig. 7. ROC curves obtained by using the image compression detection technique in [22] to identify JPEG 2000 compression in a set of unaltered and JPEG 2000 compressed images (left) and a set of unaltered and compressively sensed images (right). In the right figure "false alarms" correspond only to unaltered images misclassified as JPEG 2000 compressed. Since there is no JPEG 2000 compressed image in the seconde test set, the results in the right figure demonstrate that compressive sensing can be easily misidentified as JPEG 2000 compression.

the compression history detection technique proposed in [22]. When performing this experiment, we used the Uncompressed Colour Image Database (UCID) [35] to create a testing database of 300 unaltered images, 300 JPEG 2000 compressed images, and 300 compressively sensed images. First, we evaluated the baseline performance of the wavelet-based compression detection technique from [22] by using it to distinguish between the set of unaltered and JPEG 2000 compressed images. An ROC curve showing the results of this experiment is displayed in the left figure of Fig. 7, which shows that this technique can reliably detect wavelet-based compression. Next, we used this technique to identify evidence of JPEG 2000 compression in the set of compressively sensed and unaltered images. Since none of the images in this second experiment were compressed using JPEG 2000, we would expect the detector to find no evidence of JPEG 2000 compression. An ROC curve showing the results of this experiment is displayed in the right figure of Fig. 7. "false alarms" correspond only to unaltered images misclassified as JPEG 2000 compressed, and "detections" correspond to compressively sensed images been identified as JPEG 2000 compressed images. These results show that compressively sensed images can be easily misidentified as images that have undergone JPEG 2000 compression by existing forensic techniques. This reinforces the need for a technique to distinguish between compressive sensing and traditional wavelet-based compression.

Moreover, while the proposed universal detection schemes in Section IV can be used on images to distinguish compressively sensed images from traditionally sensed images, their performance may be affected when traditionally sensed but wavelet-based compressed images are involved in the acquisition detection analysis. To demonstrate this, we used the universal detector proposed in section IV-B to differentiate between compressively sensed images and both uncompressed traditionally

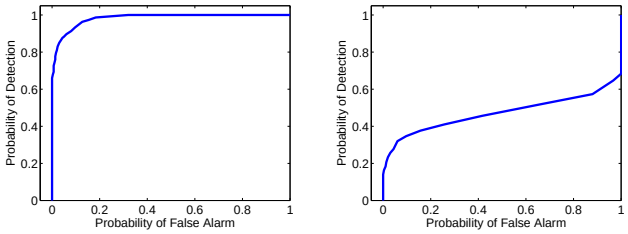


Fig. 8. ROC curves obtained by using the proposed scheme in Section IV-B to identify compressively sensed images from traditionally sensed images (left) and to identify compressively sensed images from traditionally sensed but JPEG 2000 compressed images (right).

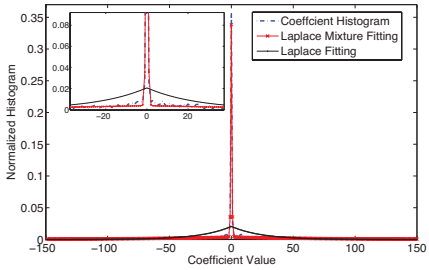


Fig. 9. Fit the coefficient histogram of compressively sensed Lena with both Laplace model and Laplace mixture model. Coefficients are taken from the third subband after 6-level DWT decomposition with wavelet basis 'bior4.4'.

sensed images as well as traditionally sensed images that have been compressed using JPEG 2000. The results of this experiment are shown in Fig. 8. The left figure demonstrates that our proposed general compressive sensing detection scheme can be successfully used on image signals. While the right figure shows the degradation of this scheme's performance when traditionally sensed but JPEG 2000 compressed images are involved in the analysis. Therefore, in order to determine the acquisition process of an image signal and identify compressive sensing, we need more specific models for compressively sensed images to distinguish them from traditionally sensed but wavelet-based compressed images.

B. DWT Coefficient Distribution Models

Because both compressive sensing fingerprints and wavelet-based compression fingerprints present themselves in an image's DWT coefficient distribution, we must adopt a set of models for an image's DWT coefficient distribution in order to develop our forensic technique. Let X be a random variable representing the value of a DWT coefficient in a particular subband of an image. For uncompressed images, we model the distribution of X using the Laplace distribution [36]

$$f_X(x) = \frac{\lambda_0}{2} e^{-\lambda_0 |x|}. \tag{19}$$

Since traditional DWT-based image compression is equivalent to nonuniform quantization [36], we then model the DWT coefficient distribution of an image that has undergone traditional wavelet-based compression as

$$\mathbf{P}[X = q] = \int_{q-\triangle_q}^{q+\triangle_q} \frac{\lambda_0}{2} e^{-\lambda_0 |x|} dx, \tag{20}$$

where $q \in \mathbb{Z}$ and $\triangle_q$ is half of the width of the quantization interval that maps DWT coefficients to q.

When examining compressively sensed images, we must account for the noise introduced into the image's DWT coefficients described in Section V-A. Since this noise will slightly smooth out the impulsive spike that we would expect to occur in the distribution of X at zero, we instead model the DWT coefficients of a compressively sensed image using a Laplace mixture distribution [37]

$$f_X(x) = \omega_1 \frac{\lambda_1}{2} e^{-\lambda_1 |x|} + \omega_2 \frac{\lambda_2}{2} e^{-\lambda_2 |x|} \tag{21}$$

where $\omega_1 + \omega_2 = 1$ and $0 < \lambda_1 < 1 < \lambda_2$. Fig. 9 shows an example of a compressively sensed image's DWT coefficient histogram fit to both a Laplace and a Laplace mixture distribution. We can see from this figure that an appropriately chosen Laplace mixture distribution very accurately models the compressively sensed image's DWT coefficient distribution.

C. Compressive Sensing Detection

Because the fingerprints left by traditional wavelet-based compression techniques can be confused with the compressive sensing fingerprints, we propose performing compressive sensing detection on images in two steps [37]. In the first step, we separate unaltered traditionally sensed images from those that are either traditionally compressed or compressively sensed. In the second step, we differentiate between compressively sensed images and those that have traditionally undergone wavelet-based compression.

Step 1 - Identify Uncompressed Traditionally Sensed Images

The goal of the first step of our compressive sensing detection scheme is to remove uncompressed traditionally sensed images from further examination. This step is equivalent to differentiating between the following two hypotheses

$$\begin{aligned} &H_0\text{: The image is uncompressed and traditionally sensed,} \\ &H_1\text{: The image is traditionally compressed} \\ &\qquad \text{or compressively sensed.} \end{aligned} \tag{22}$$

where hypothesis H_1 is a composite hypothesis. To accomplish this, we exploit the fact that the DWT coefficient distributions of both compressively sensed images and traditionally compressed images will significantly differ from the Laplace distribution.

We begin by assuming that hypothesis H_0 is correct. Under this assumption, the parameter λ_0 in (19) can be estimated for a particular subband of an image's DWT coefficients using the maximum likelihood estimator

$$\hat{\lambda}_0 = \frac{N}{\sum_{i=1}^{N} |x_i|}, \tag{23}$$

where each x_i represents a DWT coefficient in the subband being examined and N is the number of DWT coefficients in the subband. Once the estimate $\hat{\lambda}_0$ is obtained, we use $\hat{\lambda}_0$ and (19) to calculate the expected histogram g_k^{unalt} according to (14). We then measure the mean squared distance (MSD) between the observed histogram of DWT coefficients h_k and g_k^{unalt} according to the formula

$$MSD_1 = \frac{1}{B} \sum_k \left(h_k - g_k^{unalt}\right)^2, \tag{24}$$

where B is the total number of histogram bins.

We note that this step is an application of our distribution-based detection scheme proposed in section IV-B. MSD is chosen instead of KL divergence to avoid the "divide by zero" problem when calculating the KL divergence.

If the MSD between h_k and g_k^{unalt} is sufficiently large, we conclude that an image's DWT coefficient histogram cannot be modeled using (19), therefore the image either has undergone wavelet-based compression or has been compressively sensed. As a result, we differentiate between the hypotheses in (22) using the decision rule

$$\delta_1 = \begin{cases} H_0: \text{ If } MSD_1 < \tau_1 \\ H_1: \text{ If } MSD_1 \geq \tau_1, \end{cases} \quad (25)$$

where τ_1 is the decision threshold. If δ_1 returns a decision of H_1 for an image, then we proceed to step 2 of our detection process.

Step 2 - Detect Compressive Sensing

Once we have decided that an image has been either traditionally compressed or compressively sensed, we must differentiate between these two possibilities. In the second step of our detector, we frame this problem as deciding between the hypotheses

$$\begin{matrix} H_0\text{:The image has undergone wavelet-based compression,} \\ H_1\text{:The image was compressively sensed.} \end{matrix} \quad (26)$$

We know that under hypothesis H_1, an image's DWT coefficient distribution will be given by (21). As a result, we can identify compressively sensed images by determining how well the distribution of an image's DWT coefficients within a subband fits a Laplace mixture distribution.

To do this, we first estimate the parameters in the parameter set $\theta = \{\omega_1, \omega_2, \lambda_1, \lambda_2\}$ using the expectation maximization (EM) algorithm [38]. Let Z_i be latent random variables that denote which component of the Laplace mixture distribution each DWT coefficient x_i originates. As a result, we can write the following equations:

$$f_{X_i}(x_i|Z_i = 1) = \frac{\lambda_1}{2} e^{-\lambda_1 |x_i|}, \quad (27)$$

$$f_{X_i}(x_i|Z_i = 2) = \frac{\lambda_2}{2} e^{-\lambda_2 |x_i|}, \quad (28)$$

$$\mathbf{P}[Z_i = 1] = \omega_1 \text{ and } \mathbf{P}[Z_i = 2] = \omega_2. \quad (29)$$

At the t^{th} iteration of the EM algorithm, the updated estimates of the parameters in the parameter set are given by the equations

$$w_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^{N} T_{j,i}^{(t)} \quad j = 1, 2 \quad (30)$$

$$\lambda_j^{(t+1)} = \frac{\sum_{i=1}^{N} T_{j,i}^{(t)}}{\sum_{i=1}^{N} T_{j,i}^{(t)} |x_i|} \quad j = 1, 2 \quad (31)$$

where

$$T_{j,i}^{(t)} = \frac{w_j^{(t)} \lambda_j^{(t)} e^{-\lambda_j^{(t)} |x_i|}}{\omega_1^{(t)} \lambda_1^{(t)} e^{-\lambda_1^{(t)} |x_i|} + \omega_2^{(t)} \lambda_2^{(t)} e^{-\lambda_2^{(t)} |x_i|}}. \quad (32)$$

The EM algorithm's iterations are terminated after either the maximized log-likelihood ratio

$$\begin{aligned} &\max_{\theta} Q(\theta|\theta^{(t)}) \quad (33) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{2} T_{j,i}^{(t)} \left[\ln \left(\omega_j^{(t+1)} \lambda_j^{(t+1)} / 2 \right) - \lambda_j^{(t+1)} |x_i| \right]. \end{aligned}$$

converges or a fixed number of iterations have been reached.

After the values of $\omega_1, \omega_2, \lambda_1$, and λ_2 have been estimated, we compute the expected DWT coefficient histogram g_k^{cs} under hypothesis H_1 using (14). Next, we calculate the MSD between the g_k^{cs} and the observed histogram of DWT coefficients h_k

$$MSD_2 = \frac{1}{B} \sum_{k} \left(h_k - g_k^{cs} \right)^2, \quad (34)$$

where B is the total number of histogram bins. Finally, we perform compressive sensing detection according to the decision rule

$$\delta_2 = \begin{cases} H_0: \text{ If } MSD_2 > \tau_2 \\ H_1: \text{ If } MSD_2 \leq \tau_2, \end{cases} \quad (35)$$

where τ_2 is a decision threshold.

VI. Measurement Number Estimation

Once a signal has been identified as compressively sensed, a forensic investigator may wish to ascertain additional information about how the signal was captured. One significant piece of information is the number of compressive measurements that were used to acquire the signal. In this section, we propose a technique to estimate the number of compressive measurements obtained when sensing a signal.

When a compressively sensed signal is reconstructed by solving (3), the sparsest solution $\underline{x}$ such that $\Phi \underline{x} = \underline{y}$ is chosen. Since the values of $\underline{x}$ can be thought of as weights for the column vectors of Φ, and $\underline{y}$ is obtained also by weighted sum of these vectors with non-sparse weighting values, it seems natural that the sparsity of the reconstructed signal will be closely related to dimension of the column vectors of Φ, which is approximated to be the rank of Φ, i.e., the number of compressive measurements. In fact, we are able to prove that the relationship between the number of compressive measurements and the number of zeros in the reconstructed signal is given by the relationship stated below in Theorem 1.

Theorem 1: Let $\underline{y}$ be a vector of m compressive measurements obtained by compressively sensing a signal that fits one of the three signal models proposed in Section II-B. Assume that the noise, if applicable, is continuously distributed. Additionally, let the m by n sensing matrix Φ have orthonormal row vectors selected uniformly at random from an orthonormal vector set in $\mathbb{R}^n$. If the reconstructed signal $\underline{x}$ is obtained by solving the l_1 minimization problem

$$\min_{\underline{\tilde{x}}} ||\underline{\tilde{x}}||_{l_1} \quad s.t. \;\; \Phi \underline{\tilde{x}} = \underline{y}, \quad (36)$$

then with probability close to one, $\underline{x}$ will have m non-zero coefficients. As a result, the number of compressive measurements is given by

$$m = n - \Lambda_0(\underline{x}), \quad (37)$$

where $\Lambda_0(\underline{x})$ denotes the number of zero valued entries in $\underline{x}$.

Proof: We prove this theorem by deriving a lower and upper bound on $n - \Lambda_0(\underline{x})$ respectively, then showing that the only value of $n - \Lambda_0(\underline{x})$ that satisfies both bounds is m.

To derive the lower bound, we begin by defining vector space V as the linear span of the column vectors $\underline{\phi}_1, \underline{\phi}_2, \ldots, \underline{\phi}_n$ of the sensing matrix Φ. Since Φ has orthogonal row vectors, it is full rank. Thus, $\dim(V) = \dim\{\underline{\phi}_1, \underline{\phi}_2, \ldots, \underline{\phi}_n\} = m$. Next, we define the dimension of an m length vector $\underline{v}$ on space V as the size of the smallest subset of $\{\underline{\phi}_1, \underline{\phi}_2, \ldots, \underline{\phi}_n\}$ whose linear span contains $\underline{v}$.

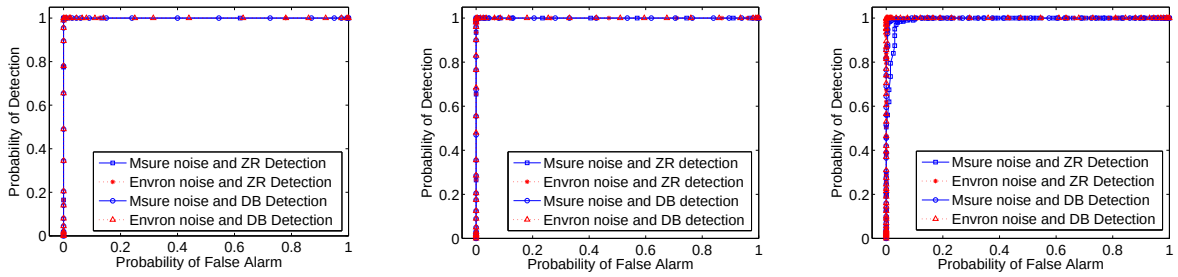
The compressive measurements $\underline{y}$ can be expressed as $\underline{y} = \sum_{i=1}^{n} \underline{\phi}_i s_i$, where $\underline{s}$ is the signal being acquired by compressive sensing. If $\underline{s}$ fits any of the signal models in Section II-B, then the dimension of $\underline{y}$ is equal to the dimension of V with probability close to one. Specifically, in the case of signals corrupted by environmental noise and nearly sparse signals, either the noise or the nature of the signal itself will cause each entry of $\underline{s}$ nonzero. Otherwise, if the signal is corrupted by measurement noise, then the independent white noise added to the compressive measurements will cause $\underline{y}$ to lie in the span of any subset of V of size $m-1$ or less with probability nearly zero.

Because the reconstructed signal $\underline{x}$ is just another decomposition of $\underline{y}$ on space V, the number of non-zero entries in $\underline{x}$ can not be less than the dimension of $\underline{y}$ on this space. Thus,

$$n - \Lambda_0(\underline{x}) \geq \dim(\underline{y}) = \dim(V) = m. \quad (38)$$

To derive the upper bound, we reformulate (36) as the following equivalent problem [25]

$$\min_{\underline{\tilde{z}}} \underline{1}^T \underline{\tilde{z}}, \quad s.t. \;\; A\underline{\tilde{z}} = \underline{y}, \;\; \underline{\tilde{z}} \geq 0. \quad (39)$$



(a) (b) (c)

Fig. 10. ROC curves of zero ratio detector and distribution-based detector on signals modeled as sparse signals in the presence of noise for (a) M/N=0.1, (b) M/N=0.4 and (c) M/N=0.9. 'Msure' is short for measurement and 'Envron' is short for environment. 'ZR' denotes the zero ratio detector and 'DB' denotes the distribution-based detector.

where $\underline{1}$ denotes a column vector of length $2n$ of all ones and $A = (\Phi, -\Phi)$ is of size $m \times 2n$. If the solution to (39) is partitioned into two vectors of equal length such that $\underline{z} = (\underline{u}^T, \underline{v}^T)^T$, then the solution to (36) can be expressed as $\underline{x} = \underline{u} - \underline{v}$.

By examining this intermediate problem, the following lemma and corollary can be proved by using Karush-Kuhn-Tucker conditions [39].

Lemma 1: Let $\underline{z}'$ denote the sparsest solution of problem (39), i.e., the one with smallest number of non-zero coefficients. Then

$$n - \Lambda_0(\underline{z}') \leq m. \tag{40}$$

Corollary 1: For any solution $\underline{z}$ of (39), the corresponding solution $\underline{x}$ for (36) will have the same number of non-zero coefficients with $\underline{z}$.

Given these two results, we conclude our proof by recalling that the solution to (36) is unique (see *Theorem 1.1* in [8]), so that the sparsest solution $\underline{x}'$ to (36) is the only solution, i.e., $\underline{x} = \underline{x}'$. Therefore, $n - \Lambda_0(\underline{x}) = n - \Lambda_0(\underline{z}') \leq m$. Combining this result with (38), we conclude that $n - \Lambda_0(\underline{x}) = m$, thus Theorem 1 is proved. ■

In practice, a number of iterative techniques are often used to solve (36). Since these techniques are typically terminated after the difference between two iterations is sufficiently small or a fixed number of iterations has been reached, the solution yielded by these techniques will often differ slightly from the optimal solution. As a result, several values of $\underline{x}$ that would ideally be zero will instead take small nonzero values. To compensate for this effect, we instead count the number of entries $\Lambda_\zeta(\underline{x})$ that fall within a ball of radius ζ around zero. Our measurement number estimator for the observed signal $\underline{x}$ is defined as follows:

$$\hat{m} = n - \Lambda_\zeta(\underline{x}), \tag{41}$$

where $\zeta = ||\breve{\underline{x}}||_\infty / \rho$. If the signal $\underline{x}$ is modeled as a sparse signal in noise, $\breve{\underline{x}}$ is taken as the noise component, otherwise $\breve{\underline{x}} = \underline{x}$. The choice of ρ depends on how accurate the reconstruction is. For example, in the ideal where the iteration in simulation can go to infinity, then $\rho \to \infty$ and $\zeta \to 0$. In our simulations, we have experimentall observed that $\rho = 100$ yields desirable performance.

VII. Simulations and Results

To verify the effectiveness of our proposed forensic techniques, we have evaluated their performance through a series of experiments. In this section, we present the results of these experiments and show that our proposed techniques can reliably detect the use of compressive sensing. We first evaluate the ability of our forensic techniques to identify compressive sensing in sparse signals in the presence of noise, nearly sparse signals, and nearly sparse signals in the presence of noise. We then evaluate the performance of our compressive sensing detection technique for images and our technique to estimate the number of compressive measurements used to acquire a signal.

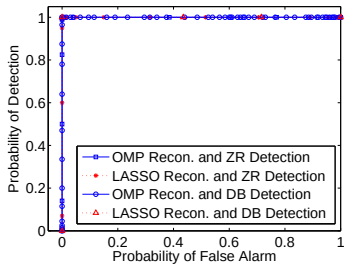
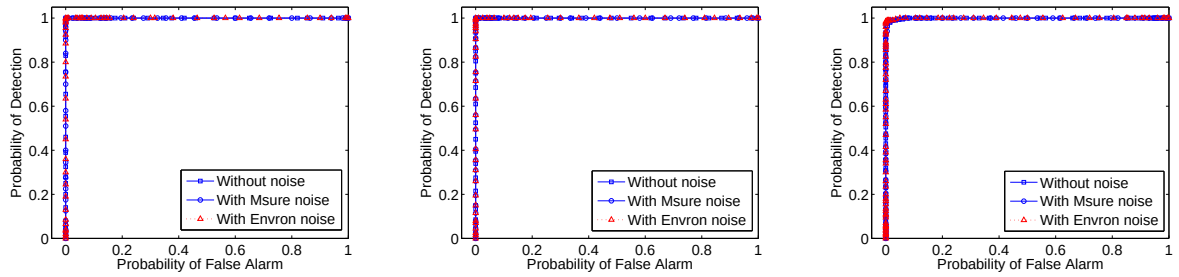


Fig. 11. ROC curves of zero ratio detector and distribution-based detector on signals modeled as sparse signals in the presence of noise when different reconstruction algorithms were used.

A. Sparse Signals in the Presence of Noise

To evaluate the ability of both the zero ratio detector and the distribution-based detector to identify compressive sensing in sparse signals in the presence of noise, we first created a database of testing signals. This database consisted of 200 compressively sensed sparse signals in the presence of environmental noise, 200 compressively sensed sparse signals in the presence of measurement noise, and 200 sparse signals in the presence of additive noise which were not compressively sensed. Each signal was created by first randomly generating a sparse signal of length $N = 1000$ with 20 nonzero entries. For each nonzero entry, its location was chosen uniformly at random and its value was drawn from a Gaussian distribution with a mean of 10 and unit variance. We then corrupted each signal with additive Gaussian noise distributed $\mathcal{N}(0, 0.1)$. For signals which were not compressively sensed, we added the noise directly to the sparse signal to obtain the observed signal. For compressively sensed signals corrupted by environmental noise, we added the noise to the sparse signal, then performed M compressive measurements. For signals corrupted by measurement noise, we first obtained M compressive measurements of the sparse signal, then added the Gaussian noise to each compressive measurement. Each compressively sensed signal was reconstructed using the basis pursuit de-noising algorithm [25]. We obtain the noise component of the observed signal by excluding the 20 entries that have the largest magnitudes, since these likely correspond to the nonzero components of the sparse signal. We then used both detection techniques to determine if each signal was compressively sensed.

In our first set of experiments, we evaluated the performance of both detection techniques as the ratio of the number of compressive measurements to the total signal length was varied from $M/N = 0.1$ to 0.9 in increments of 0.1. For distribution-based detector, the modified KL divergence was chosen as the



(a) (b) (c)

Fig. 12. ROC curves of distribution-based detection on nearly sparse signals and nearly sparse signals in the presence of noise for (a) M/N=0.1, (b) M/N=0.4 and (c) M/N=0.9. 'Msure' is short for measurement and 'Envron' is short for environment.

distance measure for it performs better than other distance measures do. When performing these experiments, we varied the decision thresholds of each detector over a range of values. For each threshold value, we determined the associated probabilities of detection P_d and false alarm P_f by calculating the percentage of compressively sensed signals that were correctly identified and the percentage of signals that were incorrectly identified as compressively sensed respectively. We then used these probabilities to construct a set of ROC curves showing the performance of each detector. Selected ROC curves showing the performance of both detectors for $M/N = 0.1$, 0.4, and 0.9 are shown in Fig.s 10(a) through (c).

From the full set of ROC curves, we found that both detectors achieved perfect detection, i.e. $P_d = 100\%$ with $P_f = 0\%$, for $M/N \leq 0.8$. When M/N reaches 0.9, both detectors can still identify compressive sensing with $P_d = 99\%$ at a $P_f \leq 5\%$. Since in most real world scenarios compressive sensing will be applied with M/N less than 0.5, these results show that both techniques perform strongly under realistic conditions. Furthermore, we can see from Fig. 10(c) that the distribution-based detector outperforms the zero ratio detector because the forensic investigator is able to make use of additional information about the noise's distribution. We also note that the performance of our detectors decrease as M increases because with more compressive measurements, the noise can be accurately reconstructed. Since compressive sensing fingerprints manifest themselves as changes in the noise distribution, this impedes compressive sensing detection. Nevertheless, our results show that compressive sensing detection can be performed with a high degree of accuracy under realistic values of M/N.

Next, we evaluated the robustness of both detectors to different signal and noise powers, as well as different noise distributions. To evaluate the performance with different signal and noise powers, we fixed the number of compressive measurements so that $M/N = 0.5$. This was done because $M/N = 0.5$ is typically an upper bound in real world applications [9], therefore it provides a lower bound on the performance of both detectors in realistic scenarios. We then repeated the previous experiments using the same noise power with signal powers of 10, 100, and 1000, and while using the same signal power with noise powers of 0.1, 1, and 10. For each of these experiments, both detectors achieved $P_d = 100\%$ at a false alarm rate of $P_f = 0\%$. These results show that both detectors can perform strongly under a variety of signal and noise powers. Next, we kept $M/N = 0.5$ and performed compressive sensing detection when each signal corrupted by noise from the exponential, Laplace, Gaussian, uniform and Rayleigh distributions. Again, under each scenario both detectors were able to acheive $P_d = 100\%$ at a false alarm rate of $P_f = 0\%$. Taken together with our previous results, these results show that both our zero ratio detector and distribution-based detector can be used to identify compressive sensing in sparse signals corrupted by noise under a wide range of conditions.

In addition, since several different algorithms are available to reconstruct a compressively sensed signal, we performed a set of experiments to demonstrate the robustness of our compressive sensing detection technique to different reconstruction algorithms. In these experiments, we used used both orthogonal matching pursuit (OMP) [30] and the LASSO error variation minimization reconstruction algorithm [31] to reconstruct the compressively sensed signals. We then repeated our first set experiments, this time setting $M/N = 0.5$. ROC curves obtained from the results of these experiments are shown in Fig. 11. These results demonstrate that both of our detectors can identify compressive sensing regardless of the reconstruction algorithm.

B. Nearly Sparse Signals and Nearly Sparse Signals in the Presence of Noise

For nearly sparse signals and nearly sparse signals in the presence of noise, we evaluated our distribution-based detector's ability to identify compressive sensing. To do this we created a testing database of 1000 signals consisting of 200 of each of the following types of signals; compressively sensed nearly sparse signals, nearly sparse signals which were not compressively sensed, compressively sensed nearly sparse signals corrupted by environmental noise, compressively sensed nearly sparse signals corrupted by measurement noise, and nearly sparse signals corrupted by additive noise which were not compressively sensed.

Each signal was generated by first creating a nearly sparse signal of length $N = 1000$ whose entries were drawn from a Laplace distribution with variance 10^4. The Laplace distribution was chosen because it is commonly used to model the coefficients of several nearly sparse signals [40], [41]. For compressively sensed nearly sparse signals, we performed M compressive measurements of the signal, then reconstructed it. For compressively sensed nearly sparse signals in the presence of noise, we applied zero mean additive Gaussian noise with variance 10 to either the signal or the M compressive measurements, then performed reconstruction using the basis pursuit denoising algorithm. To create nearly sparse signals in noise which were not compressively sensed, we added zero mean Gaussian noise with variance 10 to the nearly sparse signal. We then used our distribution-based detector to determine if each signal had been compressively sensed.

In our first set of experiments on these signals, we varied the ratio of the number of compressive measurements to the signal length from $M/N = 0.1$ to 0.9 in steps of 0.1 as was done in Section VII-A. We evaluated our distribution-based detector's performance by varying its decision threshold over a range of values, calculating the corresponding P_d and P_f for each

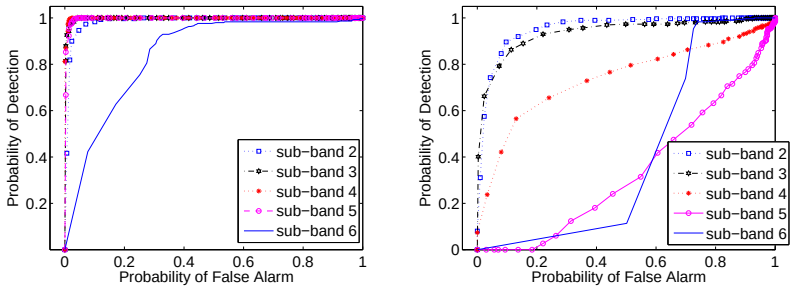


Fig. 13. ROC curves of the first (left) and second (right) step detections on each DWT sub-band coefficients. $M/N = 0.25$ is used in compressive sensing.

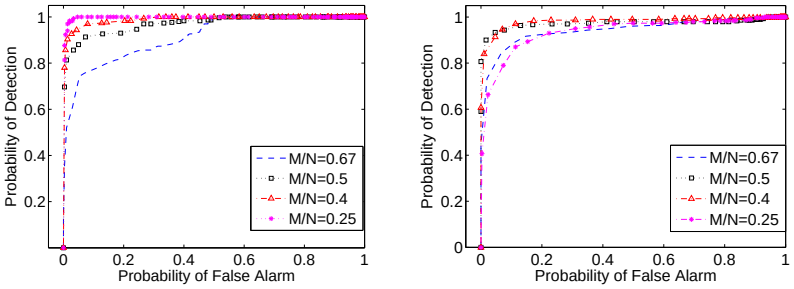


Fig. 14. ROC curves of the first (left) and second (right) step detections on coefficients of DWT sub-band 3 under different compression ratios of compressive sensing.

threshold value, then creating a set of ROC curves. Selected ROC curves for $M/N = 0.1$, 0.4, and 0.9 are shown in Fig. 12.

From the full set of ROC curves we found that when $M/N \leq 0.8$, our distribution-based detector could achieve a probability of detection of $P_d = 100\%$ with $P_f = 0\%$ for both nearly sparse signals and nearly sparse signals in the presence of either type of noise. When M/N was increased to 0.9, our detector was able to achieve a performance of $P_d = 99\%$ with $P_f \leq 3\%$ for all cases. These results show that our distribution-based detector can accuratley identify compressively sensed nearly sparse signals and nearly sparse signals in noise for realistic values of M/N.

Next, we evaluated our distribution-based detector's robustness when performing compressive sensing detection on nearly sparse signals and nearly sparse signals in noise. To do this, we performed a series of experiments in which we fixed M/N at 0.5 as was done in Section VII-A, then varied the signal variance as well as the noise power and distribution when appropriate. For nearly sparse signals, we allowed the signal variance to take values of 10^{-4}, 1 and 10^4. In each case, the detector achieved $P_d = 100\%$ with $P_f = 0\%$, i.e. perfect detection. For nearly sparse signals in the presence of noise, we repeated experiments using signal powers of 10^3, 10^4 and 10^5 and with noise powers of 0.1, 1 and 10. Additionally, we performed experiments in which we fixed the signal power at 10 and varied the noise distribution between the Gaussian, Rayleigh, Laplace, exponential and uniform distributions. In each of these experiments, our detector was able to achieve $P_d = 100\%$ with $P_f = 0\%$. These results show that our detector can be used to reliably identfy compressive sensing in both nearly sparse signals and nearly sparse signals in the presence of noise under a wide variety of conditions.

C. Images

To evaluate the performance of our compressive sensing detection technique for images, we first created a testing database of images. For each experiment, we used 300 unaltered images, 300 JPEG 2000 compressed images, and 300 compressively sensed images from the UCID database [35]. Each image in this database has size of 512×256 pixels. During JPEG 2000 compression and compressive sensing reconstruction, the 'bior4.4' DWT basis was used to perform the discrete wavelet transform of each image. To fairly evaluate our detector, during each set of experiments the compression quality factor for the JPEG 2000 images and the number of compressive measurements for the compressively sensed images were chosen so that both sets of images had the same average PSNR. For example, the average PSNRs for $M/N = 0.67$ and $M/N = 0.25$ are 36dB and 26dB, respectively.

In our first experiment, when performing compressive sensing we chose the compression ratio to be $N/M = 4$. After creating an appropriately compressed set of JPEG 2000 images, we classified each image in the testing database using our two-step image compressive sensing detection technique. When doing this, we obtained classification results using DWT subbands 2 through 6 for both detection steps. We used these results to create the set of ROC curves for each step of our detection scheme shown in Fig. 13.

The leftmost plot in Fig. 13 shows ROC curves for the first step of our detection process in which unaltered images are separated from both JPEG 2000 compressed and compressively sensed images. From these results, we can see that performing detection on subbands 3, 4, or 5 yields the best performance. For each of these subbands, our detector achieves a P_d of 100% at a P_f of 4% or less. The rightmost plot in Fig. 13 shows ROC curves for the second step of our detector. From these curves we can see that when using subbands 2 or 3 to perform detection, our detector achieves a P_d of approximately 90% at $P_f = 10\%$. Taken together, these results show that the detection scheme proposed in Section V-C can be used to reliably discriminate between unaltered, compressively sensed, and JPEG 2000 compressed images. For both steps of the detection process, we note that the performance decreases sharply when subband 6 or higher is used to perform detection. This is because the kurtosis of the distribution of DWT coefficients typically increases as the subband increases. This, together with the fact that the effective quantization interval used in JPEG 2000 is typically larger for higher DWT subbands, will result in the DWT coefficient distributions of unaltered, compressively sensed, and JPEG 2000 compressed images appearing very similar.

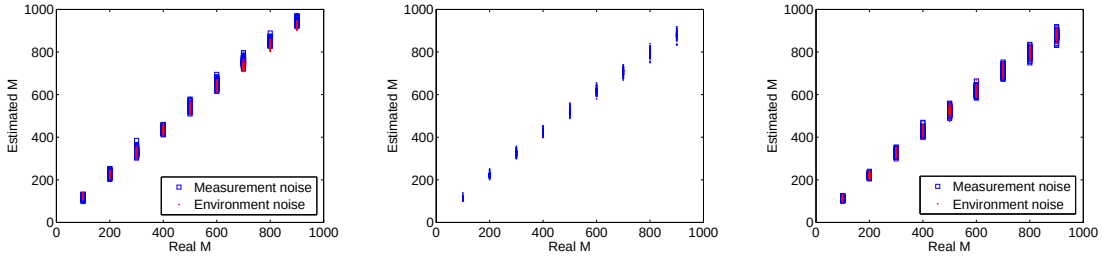


Fig. 15. Estimated $\hat{M}$ versus the real M for (a) sparse signals in the presence of noise, (b) nearly sparse signals, (c) nearly sparse signals in the presence of noise.

Next, we repeated the previous experiment while varying the number of compressive measurements so that the compression ratio of the compressively sensed images ranged between $N/M = 1.5$ and 4. In this set of experiments, we used subband 3 to perform both steps of our detection process. We used the results of this set of experiments to create the ROC curves shown in Fig. 14. We can see from the leftmost plot in Fig. 14 that the first step of our detector can achieve $P_d > 90\%$ with $P_f < 5\%$ when $N/M \geq 2$. Since in most realistic scenarios $N/M > 2$, these ROC curves show that the first step of our detector performs strongly. The rightmost plot in Fig. 14 shows that the second step of our detector can acheive a P_d of approximately 90% or higher at $P_f = 10\%$ for each value of N/M. These results show that our dector can be used to reliably discriminate between unaltered, compressively sensed, and JPEG 2000 compressed images in a variety of scenarios.

D. Estimator of the Number of Compressive Measurements

We performed a final set of experiments to evaluate the performance of our technique to estimate the number of compressive measurements used to capture a signal. In these experiments, we created a set of sparse and nearly sparse signals of length $N = 1000$ as was done in Sections VII-B and VII-A, then corrupted them using both environmental and measurement noise to create a database of 100 of each of the following signals; sparse signals in the presence of environmental noise, sparse signals in the presence of measurement noise, nearly sparse signals, nearly sparse signals in the presence of environmental noise, and nearly sparse signals in the presence of measurement noise. When creating signals corrupted by noise, we used Gaussian noise whose variance corresponded to a signal to noise ratio (SNR) of 10^3. This was done because the performance of our forensic technique decreases as the SNR decreases, thus our results can be interpreted as a conservative evaluation of our estimator's performance.

Once we created our testing database, we compressively sensed each signal while varying the number of compressive measurements from $M = 100$ to 900. We then used our forensic technique to obtain an estimate $\hat{M}$ of the number of compressive measurements used to acquire each signal. The results of this experiment are displayed in Fig. 15 which shows a series of plots comparing the estimated number of compressive measurements to the true number. We can see from this figure that for each signal model, our estimate closely matched the true number of measurements. Furthermore, we can see that our estimate lies within ± 25 measurements of the true number of measurements.

Additionally, we also testified the effectiveness of our proposed estimator of the number of compressive measurements on images. In these experiments, we tested our estimator on the database of compressively sensed images created in section VII-C. We have found that higher frequency subbands tend to have higher estimation accuracies due to their sufficient numbers of coefficients. Thus, we used subband 6 to estimate the number of compressive measurements in this subband, and then obtain the estimated ratio of M/N. The relative square error of the estimated M/N ratio was calculated as $\mathbb{E}\left[\frac{\left(\frac{\hat{M}}{N}-\frac{M}{N}\right)^2}{(\frac{M}{N})^2}\right]$. Table I lists these relative estimation errors for some typical choices of M/N ratios. The results show that the relative square error of our estimator on images is no greater than 5.2% for typical choices of compression ratios in compressive sensing.

TABLE I
RELATIVE ERROR OF ESTIMATING COMPRESSIVE MEASUREMENTS FOR IMAGES.

True M/N	0.25	0.4	0.5
Relative Error	3.2%	3.5%	5.2%

VIII. Conclusion

In this paper, we have proposed a set of techniques to identify the use of compressive sensing in a wide variety of signals. To do this, we first identified the fingerprints left in a compressively sensed signal. We then developed two general techniques to identify compressively sensed signals; one that operates by analyzing the ratio of zero valued entries in a signal, and another that operates by identifying changes to a signal's coefficient distribution caused by compressive sensing. Since evidence of compressive sensing in images can be confused with fingerprints left by JPEG 2000 compression, we designed a compressive sensing detection technique specifically tailored to digital images. Additionally, we proposed a technique to estimate the number of compressive measurements used to acquire a compressively sensed signal.

Our experimental results have shown that both our zero ratio and distribution-based detection schemes are able to reliably detect compressive sensing in a wide variety of realistic scenarios. Similarly, we hanve shown that our technique to identify compressive sensing in images can reliably distinguish compressively sensed images from both uncompressed and JPEG 2000 compressed images. Additionally, we have provided both a theoretical proof and experimental results verifying the effectiveness of our technique to estimate the number of compressive measurements used to acquire a signal.

References

[1] A. Swaminathan, M. Wu, and K. J. R. Liu, "Component forensics," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 38–48, Mar. 2009.

[2] A. C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, Oct. 2005.

[3] M. C. Stamm, M. Wu, and K. J. R. Liu, "Information forensics: An overview of the first decade," *IEEE Access*, vol. 1, pp. 167–200, 2013.

[4] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006.

[5] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš, "Source digital camcorder identification using sensor photo-response non-uniformity," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–25, Mar. 2009.

[6] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[7] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[8] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[9] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

[10] J. Provost and F. Lesage, "The application of compressed sensing for photo-acoustic tomography," *IEEE Transactions on Medical Image*, vol. 28, no. 4, pp. 585–594, Apr. 2009.

[11] J. Bobin, J.-L. Starck, and R. Ottensamer, "Compressed sensing in astronomy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 718–726, Oct. 2008.

[12] V. M. Patel, G. R. Easley, D. M. Healy, and R. Chellappa, "Compressed synthetic aperture radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 244 –254, Apr. 2010.

[13] E. G. Allstot, A. Y. Chen, A. M. R. Dixon, D. Gangopadhyay, and D. J. Allstot, "Compressive sampling of ECG bio-signals: Quantization noise and sparsity considerations," *Proc. IEEE BioCAS*, pp. 41–44, Nov. 2010.

[14] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 92–101, Mar. 2008.

[15] D. Wu, W.-P. Zhu, and M. N. S. Swamy, "A compressive sensing method for noise reduction of speech and audio signals," *Proc. IEEE MWSCAS*, pp. 1–4, Aug. 2011.

[16] "Bell labs invents lensless camera," *MIT Technology Review*, May 2013.

[17] T. Sun and K. Kelly, "Compressive sensing hyperspectral imager," *Proc. Computational Optical Sensing and Imaging*, p. CTuA5, Oct. 2009.

[18] D. Schneider, "New camera chip captures only what it needs," *IEEE Spectrum Magazine*, Mar. 2013.

[19] O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," *Applied Physics Letters*, vol. 95, no. 13, pp. 131110, 2009.

[20] InView Company, *http://inviewcorp.com*.

[21] C. Huo, R. Zhang, and D. Yin, "Compression technique for compressed sensing hyperspectral images," *International Journal of Remote Sensing*, vol. 33, no. 5, pp. 1586–1604, 2012.

[22] W. S. Lin, S. K. Tjoa, H. V. Zhao, and K. J. R. Liu, "Digital image source coder forensics via intrinsic fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 3, pp. 460–475, Sep. 2009.

[23] W. Luo, Y. Wang, and J. Huang, "Detection of quantization artifacts and its applications to transform encoder identification," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 810–815, Dec. 2010.

[24] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009.

[25] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[26] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, Academic Press, 3rd edition, 2008.

[27] H. G. Feichtinger, "Atomic characterizations of modulation spaces through gabor-type representations," *Rocky Mountain Journal of Mathematics*, vol. 19, no. 1, pp. 113–125, 1989.

[28] D. H. Brainard, "Hyperspectral image data," http://color.psych.upenn.edu/ /hyperspectral/.

[29] Rice Single-Pixel Camera Project, *http://dsp.rice.edu/cscamera*.

[30] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[31] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B*, vol. 58, pp. 267–288, 1996.

[32] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, pp. 1182–1195, 2007.

[33] X. Chu, M. C. Stamm, and K. J. Ray Liu, "Forensic identification of compressively sensed signals," *Proc. IEEE ICIP*, pp. 257–260, Sept. 2012.

[34] Y. L. Chen and C. T. Hsu, "Detecting doubly compressed images based on quantization noise model and image restoration," in *Proc. IEEE MMSP*, Oct. 2009, pp. 1–6.

[35] G. Schaefer and M. Stich, "UCID: an uncompressed color image database," in *Proc. SPIE*, 2004, vol. 5307, pp. 472–480.

[36] M. C. Stamm and K. J. R. Liu, "Anti-forensics of digital image compression," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1050–1065, Sep. 2011.

[37] X. Chu, M. C. Stamm, W. S. Lin, and K. J. Ray Liu, "Forensic identification of compressively sensed images," *Proc. IEEE ICASSP*, Mar. 2012.

[38] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, Nov. 1996 http://www.cs.ubc.ca/labs/scl/spgl1/.

[39] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," *Proc. 2nd Berkeley Symposium on Mathematical Statistics and Probability*, pp. 481–492, 1951.

[40] E. Y. Lam, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[41] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346 –2356, June 2008.