FORENSIC IDENTIFICATION OF COMPRESSIVELY SENSED IMAGES

Xiaoyu Chu, Matthew C. Stamm, W. Sabrina Lin and K. J. Ray Liu

Dept. of Electrical and Computer Engineering, University of Maryland, College Park

ABSTRACT

Due to the ease with which digital images can be forged, a great deal of work has been done in the field of digital image forensics. A particularly important problem is to forensically determine an image's acquisition and storage history. Recent research has shown that a new acquisition technique known as compressive sensing can be used to capture a digital image. Images acquired by compressive sensing can not be easily forensically distinguished from the images captured using standard digital cameras, nor from those which have undergone Discrete Wavelet Transform (DWT) based compressions. In this paper, we propose a two step detection scheme to identify the images acquired by compressive sensing. The first step identifies unaltered images captured using standard digital cameras and the second step separates compressively sensed images from ones compressed using DWT-based techniques. Our simulations show the first step of detection can achieve a probability of detection (P_d) of 100% for probability of false alarm (P_f) less than 4%, and the second step gets P_d nearly 90% with P_f of 10%.

Index Terms— Compressive Sensing, Digital Forensics, Image Compression.

I. INTRODUCTION

Digital images are widely used in our daily life for a variety of personal collection and official purposes. However, highly developed image editing software, such as Photoshop and Picasa, makes it easy to alter an image. As a result, the authenticity of the image is often doubted. To combat this, researchers have developed a number of digital forensic techniques. Researchers have already developed techniques to perform compression history investigation, camera model identification, identify image editing, and so on [1].

In many scenarios, an image's datapath integrity is critically important [2]. As a result, a number of forensic techniques have been designed to determine how an image was acquired and stored. Prior work has shown that the camera model used to take a picture can be identified by linking it to an estimate of the color filter array pattern and interpolation coefficients used during the image capture process [3]. The specific camera used to capture an image can be identified using its sensor pattern noise [4]. Because compression is usually performed when storing a digital image, techniques exist to detect the use of JPEG compression [5], along with other compression techniques such as SPIHT and JPEG2000 [2], [6]. Additionally, given the fact that most image editing is followed by re-compression, prior work have been done to detect double compression in both the Discrete Cosine Transform (DCT) and DWT domains [7], [8].

Recently, a new data acquisition and storage technique known as compressive sensing has been developed [9]. Compressive sensing is a technique capable of acquiring sparse data at a rate below the Nyquist rate. Its efficiency is based on the fact that a small group of non-adaptive linear projections of a compressible signal can contain enough information for reconstruction. Compressive sensing has demonstrated its success in many fields such as wireless communications, multimedia acquisition, and circuit failure analysis. The single-pixel camera is a typical application of compressive sensing in image acquisition [10]. Compared with conventional digital cameras, the single-pixel camera directly captures compressed image data rather than first sensing each pixel value, then applying compression to the resulting image. This and similar techniques will likely become increasingly important in applications where data must be acquired at high rates.

Presently, no techniques exist to identify images captured using compressive sensing. This is especially important because compressive sensing of images is often operates in the DWT domain, similar to several existing image compression techniques. Existing forensic techniques identify traditional DWT-based compression techniques by measuring the distance of an image's DWT coefficient histogram from a generalized Gaussian distribution [2] or by measuring the ratio of the number of DWT coefficients whose values fall into different intervals [6]. Experimentally, we have found that compressively sensed images exhibit properties that fool these techniques into thinking that they have been compressed using a traditional DWT-based method. Nevertheless, compressive sensing still leaves its own traces, i.e., fingerprints, which can be detected using the scheme proposed in this paper.

In this paper, we propose a new forensic technique to determine if an image was captured using compressive sensing. To do this, we first model the DWT coefficients of a compressively sensed image as the mixture of two Laplace distributions. We then show how this model differs from existing models of the DWT coefficient distributions for unaltered images and images compressed using DWT-based techniques such as SPIHT or JPEG2000. Our forensic technique is separated into two steps. The first identifies unaltered images captured using standard digital cameras. The second step separates compressively sensed images from images compressed using DWT-based techniques. We present experimental results to show the effectiveness of our proposed forensic technique.

II. FINGERPRINT OF COMPRESSIVE SENSING

In this section we show that the fingerprint of compressive sensing appears in the image's coefficient histograms. We begin by briefly reviewing the forensically significant characteristics of unaltered images captured using standard digital cameras and those that have undergone traditional DWT-based compressions.

It has been shown that the coefficient histograms in the transform domain (either DCT or DWT) are zero-mean and smooth if the image is a natural image and has not been altered in this transform domain [11]. In the past, the distribution of DWT coefficients in each sub-band has been modeled using the Laplace distribution. Fig. 1 shows an example of a typical DWT coefficient histogram from an unaltered image captured by a standard digital camera.

DWT-based compression techniques, such as SPIHT and JPEG2000, begin by decomposing an image using a multi-level wavelet transform. For each sub-band, the coefficients are quantized using bit-plane encoder so that the later transmitted data stream keeps significant bits while dropping others. When reconstructing from the receiver side, the sub-band coefficients are obtained by replacing the lost bits with zeros which causes the reconstructed coefficients to be clustered around certain integer values. This effect can also be seen in the center plot of Fig. 1, which shows the DWT coefficient histogram of an image compressed using JPEG2000.

Email: {cxygrace,mcstamm,wylin,kjrliu}@umd.edu.
This work is supported in part by AFOSR grant FA95500910179.

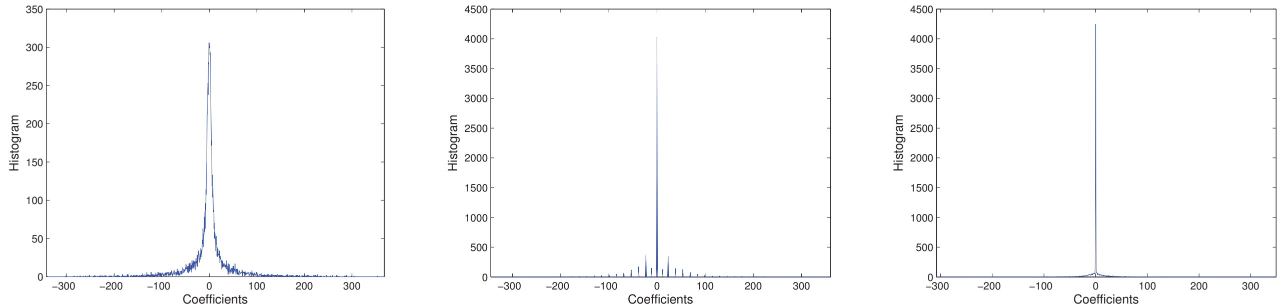

Fig. 1. Histogram of DWT coefficients taken from uncompressed *Lena* (left), the same image after JPEG2000 compression (center), and the reconstructed compressively sensed *Lena* (right).

Compressive sensing, on the other hand, is a completely different technique for image acquisition and storage. This technique measures a signal by projecting it on to random basis functions. An example of this is the single pixel camera which acquires each measurement by employing a digital micro-mirror array to optically calculate linear projections of the scene onto a specific random pattern [10]. Thus, each compressed measurement is a combination of weighted pixel values in the original image. Reconstruction can be done effectively using non-linear algorithm when the original signal has a sparse representation. Because the DWT coefficients of natural images tend to be sparse, compressive sensing on images is often performed in the wavelet domain.

The right figure in Fig. 1 shows the coefficient histogram of an image captured using compressive sensing. Unlike traditional DWT-based compression techniques which concentrate transform coefficient values into multiple peaks, this DWT coefficient distribution has one strong peak at zero. Experimentally, we have observed such distribution has significantly more kurtosis than the Laplace distribution. This happens because the reconstruction of compressive sensing retains the most significant values while throwing those small values away (set to zero). Furthermore, since this coefficient histogram is too peaked for the Laplace or generalized Gaussian distribution and also has coefficient numbers drop dramatically from duration $(0, 1)$ to $[1, 2)$, almost the same as the quantized case. Thus, [2] and [6] may fail to distinguish compressively sensed images from traditional DWT-based compressed ones.

III. DETECTION SCHEME

As we discussed in the previous section, the coefficient histogram of unaltered images taken from standard digital cameras, DWT-based compressed images and compressively sensed images have different fingerprints in transform domain. We exploit this disparity to propose our forensic detection scheme. We first introduce our model for each case, next use statistic tools to estimate the parameters of each model, then classify the test image according to the distribution that is most closely fits.

III-A. Models

We model the distribution of DWT sub-band coefficients in an unaltered image captured by a standard camera using the Laplace distribution [12]. Let random variable X denote the value of a DWT coefficient in a specific sub-band after DWT transform. The distribution of X is

$$\mathbf{P}[X = x] = \frac{1}{2\lambda_s} e^{-\frac{|x|}{\lambda_s}} \quad \text{with } \lambda_s > 0. \tag{1}$$

For images compressed using traditional DWT-based techniques, such as SPIHT or JPEG2000, we model their DWT coefficient histogram using the discrete Laplace distribution. This is because both bit plane truncation and quantization cause the DWT coefficients to cluster into peaks separated by the width of each quantization interval.

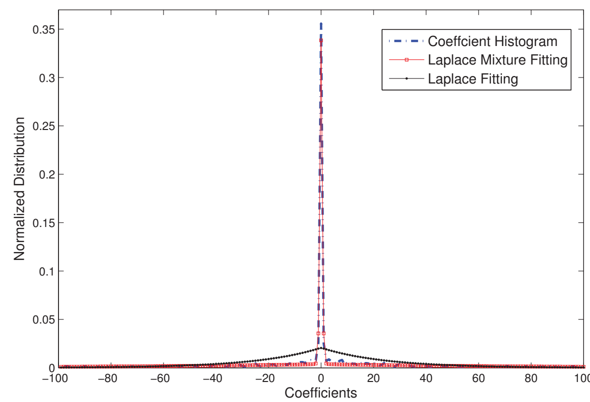

Fig. 2. Fitting the coefficient histogram of compressively sensed *Lena* with both Laplace model and Laplace mixture model.

We model the DWT domain coefficients distribution of the compressively sensed image based on experimental observations, we propose modeling the DWT using a *Laplace mixture model*. Use the same notation of coefficient variable X with (1), we have

$$\begin{aligned} \mathbf{P}[X = x] \quad &= \quad w_1 \frac{1}{2\lambda_{d1}} e^{-\frac{|x|}{\lambda_{d1}}} + w_2 \frac{1}{2\lambda_{d2}} e^{-\frac{|x|}{\lambda_{d2}}} \\ &\quad \text{with } w_1 + w_2 = 1, 0 < \lambda_{d1} < 1, \lambda_{d2} > 1. \end{aligned} \tag{2}$$

The first term approximates the central peak of the coefficient histogram, while the second one is made to fit the tails on both side. We use Laplace mixture model for it inherits the Laplace characteristic from the original image. So does the discrete Laplace model for traditional DWT-based compressed images. Fig. 2 gives an example on a particular compressively sensed image to show the effectiveness of our proposed model, where the coefficients of sub-band 3 is studied after 6-level DWT decomposition with wavelet basis bior4.4.

III-B. Detection and Classification

Based on the three models in the previous subsection, we propose a two step detection scheme, which is sketched out in Fig. 3. In particular, we use the Laplace model in (1) to first try to fit the coefficient histogram of the test image. If it fits, this image is identified as unaltered image captured by standard digital camera. If not, we proceed to the second step - fitting it into Laplace mixture model, in which fitting will identify it as an image captured by

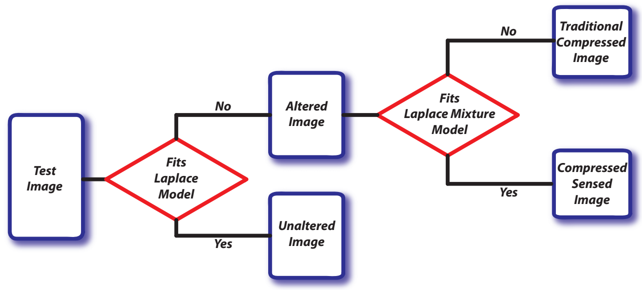

Fig. 3. Two step detection scheme to distinguish the compressively sensed image by studying the coefficient histogram in DWT domain.

compressive sensing, and otherwise it is traditional DWT-based compressed. We use two step scheme instead of only fitting the testing image into Laplace mixture model is due to the fact that Laplace distribution can certainly be presented as a combination of two Laplace distribution, while the opposite statement is not true.

First Step - Identify Unaltered Images Captured using Standard Digital Cameras

Our first detection step is to detect whether the image is unaltered and captured by a standard digital camera. The test image is first transformed into a specific DWT domain. We assume the DWT basis is known. If not, it can be found using techniques similar to those described in [2], [6]. We then estimate the parameter λ_s in the Laplace model (1) from the given data of the test image. Let $x_i, i = 1, \ldots, n$ denote the coefficients in a single sub-band, with n be the total number of coefficients in this sub-band. The estimated parameter $\hat{\lambda}_s$ in our Laplace model is obtained by the maximum likelihood estimate

$$\hat{\lambda}_s = \frac{\sum_{i=1}^{n} |x_i|}{n}.$$

Having this estimated parameter in hand, we calculate the mean square error (MSE) between the normalized observed histogram $h_o(k)$, and the histogram of estimated distribution $h_e(k)$ according to the equation

$$MSE_s = \frac{\sum_k \big(h_o(k) - h_e(k)\big)^2}{N}, \quad (3)$$

, where N denotes the total number of bins in the histogram. The estimated histogram can be obtained by integrating the distribution function within each bin range. Next we determine if the image is unaltered by performing the following threshold test

H_0 : The image is altered if $MSE_s > \tau_s$
H_1 : The image is unaltered if $MSE_s \leq \tau_s$

, where τ_s is an appropriately chosen decision threshold.

Second Step - Identify Images Acquired by Compressive Sensing

Whenever the detection result in the previous step is hypothesis H_0, we come to the second step: identify whether the test image has been captured by compressive sensing. We do this by testing to see if the coefficient histogram can be modeled as Laplace mixture distribution. We still use the coefficient data obtained at the beginning of our first detection step and estimate the parameters in (2), w_1, w_2, λ_{d1} and λ_{d2}. Expectation-maximization (EM) algorithm is used in this step [13]. This algorithm is to find the maximum likelihood estimations of parameters in statistical models, where the model depends on unobserved latent variables. In our case, given an observation, there is certain probability w_1 that it obeys the Laplace distribution with parameter λ_{d1}, yet it also may lie in the other Laplace distribution field, with parameter λ_{d2}, of probability w_2. EM is an iterative method which alternates between performing an expectation (E) step, which computes the expectation of the log-likelihood evaluated using the current estimations for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step. Such E-M iteration goes on until convergence.

In our specific case, the coefficients $x_i, i = 1, \ldots, n$ are considered as the samples of n independent observations $X_i, i = 1, \ldots, n$ from a mixture of two Laplace distributions. And let $Z_i, i = 1, \ldots, n$ be the latent variables that determine the component from which the observation originates. Then for each $i, i = 1, \ldots, n$

$$\begin{aligned} \mathbf{P}_{\lambda_{d1}}[X_i = x_i | Z_i = 1] &= \frac{1}{2\lambda_{d1}} e^{-\frac{|x_i|}{\lambda_{d1}}} \\ \mathbf{P}_{\lambda_{d2}}[X_i = x_i | Z_i = 2] &= \frac{1}{2\lambda_{d2}} e^{-\frac{|x_i|}{\lambda_{d2}}} \end{aligned}$$

$$\text{with} \quad \mathbf{P}[Z_i = 1] = w_1 \quad \text{and} \quad \mathbf{P}[Z_i = 2] = w_2.$$

The estimated parameters $\theta = \{w_1, w_2, \lambda_{d1}, \lambda_{d2}\}$ in $(t+1)^{th}$ iteration given the result of t^{th} iteration are obtained by

$$\begin{cases} w_j^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} T_{j,i}^{(t)} & j = 1, 2 \\ \lambda_{dj}^{(t+1)} = \frac{\sum_{i=1}^{n} T_{j,i}^{(t)} |x_i|}{\sum_{i=1}^{n} T_{j,i}^{(t)}} & j = 1, 2 \end{cases}$$

With $T_{j,i}^{(t)}, i = 1, \ldots, n, j = 1, 2$ defined as

$$T_{j,i}^{(t)} = \frac{w_j^{(t)} \frac{1}{2\lambda_{dj}^{(t)}} e^{-\frac{|x_i|}{\lambda_{dj}^{(t)}}}}{w_1^{(t)} \frac{1}{2\lambda_{d1}^{(t)}} e^{-\frac{|x_i|}{\lambda_{d1}^{(t)}}} + w_2^{(t)} \frac{1}{2\lambda_{d2}^{(t)}} e^{-\frac{|x_i|}{\lambda_{d2}^{(t)}}}}.$$

And the maximized log-likelihood expectation $Q(\theta|\theta^{(t)})$ according to these estimates is

$$\max_{\theta} Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n}\sum_{j=1}^{2} T_{j,i}^{(t)} \left[\ln \frac{w_j^{(t+1)}}{2\lambda_{dj}^{(t+1)}} - \frac{|x_i|}{\lambda_{dj}^{(t+1)}} \right].$$

The iteration stops whenever the maximum log-likelihood expectation converges or the limit of iteration number is reached.

Having these estimated parameters, we again calculate the MSE - MSE_d - between the normalized observed histogram and the estimated distribution using the similar equation (3). We finally determine if the image is compressively sensed by performing the following threshold test

H_0 : The image is DWT-based compressed if $MSE_d > \tau_d$
H_1 : The image is compressively sensed if $MSE_d \leq \tau_d$

, where τ_d is an appropriately chosen decision threshold.

We summarize our scheme as follows:

1) Estimate the parameter of Laplace model using the coefficients data taken from an appropriate DWT domain. Determine whether this image is unaltered and captured by a standard digital camera by calculating the MSE between observed histogram and estimated Laplace model, and see if it is below the threshold.
2) If the image is identified as altered in Step 1), use the coefficient data to estimate the parameters in Laplace mixture model. And further decide whether this image has been traditional DWT-based compressed or been captured using

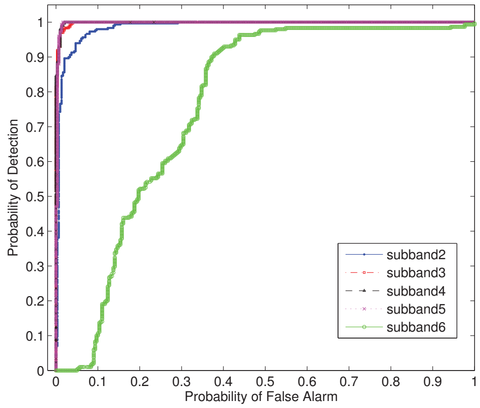

Fig. 4. ROC curve showing the performance of the first step of our detection technique. This step identifies unaltered images captured by standard digital camera and separates them from images that acquired by compressive sensing or compressed using a DWT-based technique.

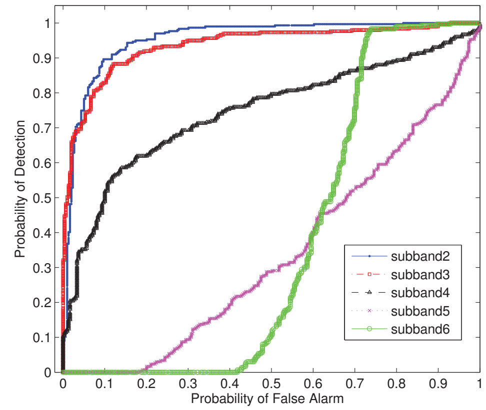

Fig. 5. ROC curve showing the performance of the second step of our detection technique. This step detects the images acquired by compressive sensing and separate them from those compressed using a DWT-based technique.

compressive sensing technique by calculating the MSE between the observed histogram and the estimated Laplace mixture model, and compare it with the threshold.

IV. SIMULATIONS AND RESULTS

To evaluate the performance of our compressive sensing detection scheme for image acquisition, we compiled a testing database of 597 images. This database contained 299 unaltered images captured using standard digital cameras, 299 images compressed using JPEG2000, and 299 compressively sensed images (i.e., measurements are random projections of the unaltered images). When compressive sensing was performed, we used the same DWT basis that was used during JPEG 2000 compression as the sparse domain for reconstruction, where spectral projected gradient pursuit algorithm was used. We then used our proposed compressive sensing detection scheme to classify each image as unaltered, compressed using a DWT based technique, or compressively sensed.

Fig. 4 plots the Receiver Operating Characteristic (ROC) curve of our first detection step to identify the unaltered images. Sub-band 2 through sub-band 6 are all the AC sub-bands in the transform domain, from lowest frequency to highest frequency. Among them, sub-band 3, 4 and 5 perform the best. When these sub-bands are used to detect unaltered images, our detector achieves a P_d of 100% for P_f less than 4%. In addition, the coefficients of the lowest frequency, sub-band 2, have more obvious noise than the others which makes it not ideal for the detection. On the other hand, the coefficients from the highest frequency have too much kurtosis that makes it harder to detect as well.

Fig. 5 shows the ROC curve to display the performance of the second detection step of our compressive sensing detection technique. Concerning the best performance sub-bands, sub-band 2 and 3, we can obtain P_d nearly 90% with P_f of 10%. This figure also shows that for the coefficients taken from higher frequencies, since the quantization step of JPEG2000 compression becoming much larger, the peaks are too small to detect, and hence it is harder to distinguish itself with the compressively sensed images.

V. CONCLUSIONS

In this paper, we proposed a technique to identify images acquired by compressive sensing and separate them from unaltered images captured using standard digital cameras and also images compressed using traditional DWT-based techniques. We proposed a model for the DWT coefficients of compressively sensed images. Our two step detection first identifies the unaltered images using maximum likelihood estimate and threshold test, then further identifies the compressively sensed image by expectation-maximization estimate and threshold test. Our simulation shows the ROC curve of each step, where we get P_d of 100% for P_f less than 4% in the first detection step, and P_d nearly 90% with P_f of 10% for the second step.

VI. REFERENCES

[1] H. Farid, "Image forgery detection," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–25, Mar. 2009.
[2] W.S. Lin, S.K. Tjoa, H.V. Zhao, and K.J.R. Liu, "Digital image source coder forensics via intrinsic fingerprints," *IEEE Trans. on Information Forensics and Security*, vol. 4, no. 3, pp. 460–475, Sep. 2009.
[3] A. Swaminathan, M. Wu, and K.J.R. Liu, "Nonintrusive component forensics of visual sensors using output images," *IEEE Trans. on Information Forensics and Security*, vol. 2, no. 1, pp. 91–106, Mar. 2007.
[4] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Trans. on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006.
[5] Z. Fan and R.L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Trans. on Image Processing*, vol. 12, no. 2, pp. 230–235, Feb. 2003.
[6] W. Luo, Y. Wang, and J. Huang, "Detection of quantization artifacts and its applications to transform encoder identification," *IEEE Trans. on Information Forensics and Security*, vol. 5, no. 4, pp. 810–815, Dec. 2010.
[7] T. Pevny and J. Fridrich, "Detection of double-compression in JPEG images for applications in steganography," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 2, pp. 247–258, Jun. 2008.
[8] J. Zhang, H. Wang, and Y. Su, "Detection of double-compression in jpeg2000 images," *2nd International Symposium on Intelligent Information Technology Application*, vol. 5, pp. 418–421, Dec. 2008.
[9] D.L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[10] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, T. Sun, K.F. Kelly, and R.G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, Mar. 2008.
[11] S.G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[12] M.C. Stamm and K.J.R. Liu, "Anti-forensics of digital image compression," *IEEE Trans. on Information Forensics and Security*, vol. 6, no. 3, pp. 1050–1065, Sep. 2011.
[13] T.K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, Nov. 1996.