IMAGE TAMPERING DETECTION USING COMPRESSIVE SENSING BASED WATERMARKING SCHEME

Mehul S. Raval*, Manjunath V. Joshi*, Priti P. Rege*, S K Parulkar*
*Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar
*College of Engineering Pune (COEP), Pune
*Computer Division, BARC, Mumbai

Abstract In this paper, we describe an image watermarking scheme using compressive sensing for detection of image tampering. We exploit dimensionality reduction and inherent computational security offered by compressive sensing. The frequency band containing the high energy is used to compute the hash and then its pseudo-random projections are obtained. These projections are embedded as a watermark into the frequency band containing low energy. On recovering the watermark, projections are acquired, which in turn are used to solve convex optimization problem to obtain the hash. The image tampering can be detected and integrity of watermark can be established by comparing the hash values.

Index Terms – Compressive sensing, convex optimization image authentication, watermarking

1. INTRODUCTION

Proving integrity and authenticity of the multimedia signal is one of the most important challenges due to penetration of digital signal into our lives. When we create a copy of the digital signal, we are actually reconstructing the same signal. This brings in simplicity offered by digital domain but introduces plethora of problem like difficulty in maintaining integrity and authenticity of digital signals. In the state of art literature three types of approaches are being followed for solving the problem of image tampering detection [1]:
1. Blind detection.
2. Image hashing
3. Using digital watermarks

Blind methods search for the correlations that exist for a specific type of attacks. The advantage with these schemes is that they do not require any side information for tampering detection [2, 3]. They are advantageous when original content is unavailable, however they can report a below par performance as they have to model each type of manipulations. Image hash is a feature based compact representation of the digital content. Image hashing has to satisfy following properties [4, 5]: 1. perceptual fidelity; 2. one way function with no reproducibility other way; 3. extremely sensitive to input data; 4. use for image quality estimation; 5. use in data base retrieval. Moreover it is combined with Compressive Sensing (CS) [6, 7] to solve issue of image authentication. CS allows reconstruction of a sparse signal from its samples, which are sampled at a rate much lower than the Nyquist rate. The main theme of CS is that small number of linear projections of sparse signals is sufficient to represent its structure faithfully. The signal can be reconstructed by solving a convex optimization problem. Digital watermark is embedded directly into digital content for image authentication and integrity check. Fragile watermarks are used to indicate loss of integrity [8, 9] when image semantics are changed. Authors in [8] used wavelet domain for embedding watermark these manipulations are localized in frequency and spatial domain. In [9] watermark is inserted in least significant bits with hierarchical fashion. This gave robustness against vector quantization attacks. Recently semi-fragile watermarking is combined with principles of CS to tackle image authentication [10, 11, 12]. Authors in [10] proposed a tampering detection, localization and recovery of original coefficients (if tampering is sparse). The embedded watermark is image dependent and method exploits the sparseness of Discrete Cosine Transform (DCT) domain. Five MSB layers of pixels are preserved and 3 LSB layers of all pixels are used for embedding watermark. Reference bits are non-uniformly quantized before embedding. The watermark extracted from the non tampered area is used for localization and content recovery is done using compressive sensing with unconstraint and constraint conditions.

Authors in [11] decimate the image and create the projections. The random projections are quantized with a uniform quantizer. Low Density Parity Check (LDPC) code is used to form the hash which is embedded as a robust watermark. Recovery of watermark leads to recovery of projections, which are in turn used to obtain estimate of the distortion in the received image. If the tampering is sufficiently sparse, localization is done. The system is based on the principles of CS and Distributed Source Coding (DSC). In [12] CS based watermark detection in transform domain (DCT) was presented. The method relies on the fact that for natural images, coefficients in transform domain are sparse. They can be further sparsified by applying threshold to coefficients. Additive watermarking scheme is used for insertion of watermark in middle frequency coefficients. Watermark using CS is recovered under certain conditions on sparsity and ratio of code word length to watermark length. The watermark robustness against Poisson noise attack has been demonstrated by the authors.

In this paper we propose a semi-fragile watermarking scheme using CS based detection. N dimensional watermark is multiplied by random Guassian matrix with size MxN to generate M dimensional projections. The random projections are embedded as a watermark and then recovered at the detector. The CS based detection offers dimensionality reduction and security. We use mean squared error (MSE) and visual information fidelity (VIF) [13, 14] measure for quantifying the tampering.
2. BACKGROUND OF COMPRESSIVE SENSING

The theory of CS [6, 7] demonstrates the recovery of sparse signal. Any signal $X$ in $\mathbf{R}^N$ can be represented as

$$X = \Psi \Phi S$$  \hspace{1cm} (1)

where $\Psi$ is an N x N basis matrix and $S$ is Nx1 column vector of weighting coefficients.

Also $S_j = \{X, \Psi_j\}$

If the signal $X$ has $K$ sparse elements, then only $K$ basis vectors are required to form to the signal. Considering measurements $y_j$ is generated as follows $y_j = \langle X, \Phi_j \rangle \cdot j = 1 \ldots M$

where $\Phi_j$ is a measurement matrix with $M < N$.

Substituting from (1) we get $y = \Phi \Psi S = AS$

where $A$ is an $MxN$ matrix and allows signal reconstruction provided $M \geq K$ and $\Phi$ and $\Psi$ must be incoherent $A$ must preserve length (restricted isometric property (RIP)) of a particular $K$ sparse vector $b$ for $\epsilon > 0$. Both RIP and incoherence is satisfied by using $\Phi$ as a Gaussian measurement matrix which has RIP with high probability and signal $X$ can be recovered provided

$$M \geq cK \log \left( \frac{N}{\epsilon^2} \right) \leq N$$  \hspace{1cm} (5)

The recovery algorithm aims to find the sparsest $S$ such that $y = AS$. This can be solved by posing it as following convex optimization problem.

$$\min ||S||_1 \text{ s.t. } y = AS$$  \hspace{1cm} (6)

This recovers exactly $K$ sparse components of the signal and problem can be converted to linear program of basis pursuit [6, 7] with computational complexity of $O(N^2)$.

3. PROPOSED WATERMARKING SCHEME

3.1 Embedding

Transform the image using Discrete Wavelet Transform and compute the hash $\tilde{W} \in \{0, 1\}$ for the approximation (LL) band. Next we encode it as sequence $y = AW$

$$A_{MN} \text{ is an orthonormal Gaussian } N(0,1/N) \text{ matrix generated from a seed } k \text{ known only to embedder and decoder. This ensures security as decoder requires same seed to generate the } A \text{ which is important in CS decoding. Observations } y \text{ are quantized using a uniform quantizer } Q \text{ and encoded output values } O \in \{0, 1\}$. These binary values are inserted using quantization index modulation (QIM) into the diagonal band (HH) band. Inverse transformation will yield watermarked version of the image. Distance between the two quantization levels can be used to control the robustness of scheme. Larger distance between two levels results in improved robustness but at the cost of degradation in perceptual quality.

3.2 Detection

Image whose integrity is to be validated is transformed using the DWT. QIM decoder will extract the encoded bits from the diagonal (HH) band. Using the quantizer module, bits are decoded and mapped to their respective quantization levels to obtain observations $\hat{y}$. Using these $M$ observations, convex optimization problem (posed by equation 6) is solved to obtain the hash $\hat{W}$ (extracted value). By applying the hash function to approximation band, hash value is recomputed ($W$). If it matches with the extracted hash value ($\hat{W} = W$), then image integrity is believed to be intact. MSE and VIF are used as a measure to quantify the results using the hash values. VIF offers robustness against the geometric transformation and exhibits high similarity between an image and its geometrically transformed versions as compared to MSE.

4. SIMULATION RESULTS

For the evaluation of the scheme, we conducted experiments on 256 x 256 gray scale images. Haar wavelet is used to split the image into its frequency constituents. We use whirlpool [15] hashing function to generate 512 bits hash. $W$ dimensions for equation 7 are as follows $y_{2090,24} = A_{2090,24}W_{532,24}$. We use seed $k$ to generate Gaussian matrix and orthonormalize it using Gram Schmidt method. Observations $y$‘s are quantized using 64 level uniform quantizer to obtain the encoded bit stream with 6 bits / level. The total number of bits for observation vector $y$ is 200 x 6 = 1200, which is embedded using QIM in diagonal band. The same quantizer is used at the receiver to reassigned level to the extracted code. We simulate three different attacks to check the fragility of the scheme, namely histogram equalization, JPEG compression (quality factor $Q=100$) and rotation along with cropping.

Peak signal to noise ratio, MSE and VIF are used as a performance metric for perceptual invisibility and fragility of watermark. Comparing the MSE between the recovered and reconstructed hash vector is equivalent to comparing MSE between the two images [14] and equivalently comparing $y$, $\hat{y}$. $W$ is treated as authentic if $MSE(y, \hat{y}) < \tau_{m\text{se}}$ where $\tau_{m\text{se}}$ represents some fixed threshold. The image authentication problem has been posed as a hypothesis testing problem in [14] to determine $\tau_{m\text{se}}$. Two hypotheses are

$$H_0: \hat{W} \text{ is authentic i.e. } V_{m\text{se}} = \text{MSE}(y, \hat{y}) < \tau_{m\text{se}}$$

$$H_1: \hat{W} \text{ is unauthentic i.e. } V_{m\text{se}} = \text{MSE}(y, \hat{y}) > \tau_{m\text{se}}$$

Considering $MSE(y, \hat{y})$ as a random variable with Gaussian i.i.d distribution and applying central limit theorem to obtain optimal value of MSE as

$$V_{m\text{se, opt}} = (b + \sqrt{(b^2 - ac)})/a$$

where $a = (\sigma_1^2 + \sigma_2^2)$, $b = (\sigma_1^2 * \mu_2 - \sigma_2^2 * \mu_1)$, and $c = (\sigma_1^2 \mu_2^2 - \sigma_2^2 \mu_1^2 - 2 \sigma_1 \sigma_2 \ln (\sigma_1 / \sigma_2))$.

$\mu_1$ and $\sigma_1$ denote the mean and standard deviation of $V_{m\text{se}}$ when $W = \hat{W}$ and $\mu_2$ and $\sigma_2$ denotes the mean and standard deviation of $V_{m\text{se}}$ when $W \neq \hat{W}$. Hence $\tau_{m\text{se}} = V_{m\text{se, opt}} > 0$.

The result of experimentation is summarized in Table I and Fig. 1 with PSNR of watermarked image as 49.61 dB and $\tau_{m\text{se}}$ as 0.0072. Table I indicates that image is declared authentic only if $MSE(y, \hat{y}) < \tau_{m\text{se}}$ all the other cases are treated as deliberate
tampering to image. Fig.1 indicates excellent perceptual quality obtained by the scheme.

Table I: Performance parameters

<table>
<thead>
<tr>
<th>Case</th>
<th>V_{mse}</th>
<th>VIF</th>
<th>Execution time</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>No tampering</td>
<td>0.0002</td>
<td>0.9</td>
<td>0.93</td>
<td>Authentic</td>
</tr>
<tr>
<td>Brightness</td>
<td>0.0455</td>
<td>0.18</td>
<td></td>
<td>Unauthentic</td>
</tr>
<tr>
<td>JPEG Q=100</td>
<td>0.0442</td>
<td>0.39</td>
<td></td>
<td>Unauthentic</td>
</tr>
<tr>
<td>Rotate 1°</td>
<td>0.1102</td>
<td>0.2</td>
<td>1.02</td>
<td>Unauthentic</td>
</tr>
</tbody>
</table>

5 CONCLUSION

In this paper we present a fragile watermarking scheme based on compressive sensing detection. The scheme has following features: 1. Blind detection; 2. computational security due to CS because without seed attacker would not get correct measurement matrix A; 3. Scheme detects the semantic changes and indicates the tampering.

Acknowledgement: This work is supported by grant from Board of Research in Nuclear Science (BRNS), DAE, BARC, Mumbai, India.

REFERENCES