A NEW SEMI-FRAGILE WATERMARKING SCHEME FOR AUTHENTICATION AND TAMPER LOCALIZATION IN REMOTE SENSING IMAGES

Salwa A.K Mostafaa, Naser El-Sheimb, A. S. Tbolba, Hisham M. Elhindya, F. M. Abdelkaderc

a Dept. of Electrical Engineering, Faculty of Engineering, Suez-Canal University, Egypt - asali@ucalgary.ca

b Dept. of Geomatics Engineering, Faculty of Engineering, University of Calgary, Calgary, Canada-elseheimy@ucalgary.ca

c Faculty of Computer Studies, Arab Open University, Kuwait, HQ - a.tolba@arabou.edu.kw


ABSTRACT:

Recently, remotely sensed satellite images are a crucial source of data for geographical information systems and are used in many applications. The process of detecting any illegitimate alteration on the content of these images becomes a new challenge for content owners and distributors. To overcome this challenge, digital watermarking techniques are developed to authenticate the remote sensing (RS) images. These techniques are based on embedding a watermark in the image to produce an authenticated version of the image so that any attempts to change the content of the authenticated image will result in a change in the watermark itself. In this paper, a new semi-fragile watermarking technique for image authentication is proposed based on integer wavelet transform (IWT). The watermark is decomposed into its bit-planes which are inserted into the bit-planes of the IWT details sub-bands of the RS image color bands. Experimental results show a good invisibility of the proposed scheme. Moreover, it is robust to some image processing attack such as JPEG compression, Gaussian noise adding and filtering, but it is sensitive to malicious attack such as cut and paste. The scheme also can detect the tempering region exactly.

I. INTRODUCTION

Nowadays, we witness the widely spreading of RS image via the internet and growing their application fields (such as agriculture, archaeology, military, homeland security... etc.). RS images are different from general images, they are high cost, high resolution images and they have complex features and precious data as military images thus the demand to copy right and authenticate these images is increased. Digital watermarking is a new effective technique that can fulfill these issues using three different categories; robust, fragile and semi-fragile watermarking.

Robust watermarking is used for copyright applications (Chrysochos et al., 2008; Reddy et al., 2009; Barni et al., 2003; Hsieh et al., 2008; Anusudha et al., 2007). It protects the copyright information into digital content against different kinds of manipulations or attack. On the other side, fragile and semi-fragile watermarking schemes are used for the purpose of authentication which is the scope of this work. Image authentication is the process of embedding watermark to verify the integrity of the image. Any malicious manipulation of the image will destroy the watermark so we can determine if the image is tampered or not and where it is tampered. Fragile watermarking is sensitive for any manipulation of the image whether this manipulation is malicious (such as content change: cut and paste) or unintentional (such as some image processing: compression, filtering ...etc.) (Devi, 2009a et al.; Hassan and Gilani 2006), while semi-fragile watermarking is at the middle between the fragile and robust watermarking. It is fragile to malicious manipulation and robust to unintentional manipulation (Seng et al., 2009).

Different techniques have been proposed for image watermark authentication but it is still limited in RS image. (Qin et al., 2004) presented a semi-fragile watermarking scheme based on wavelet packet. The edge feature of the RS image is used as a watermark and is embedded into the LL band of wavelet packet. The scheme is robust to compression and noise and it is able to localize image tampering by cutting and altering. In (Ho et al., 2004), the pseudorandom binary sequence watermark is embedded into the pinned field of the pinned sine transform (PST), which contains the texture information of the original image. The scheme is especially sensitive to texture alterations of the host image and robust to JPEG compression. (Caldelli et al., 2006), focused on authentication and tamper localization through fragile watermarking. The watermark is embedded in the Least Significant Bits (LSB) of the host image in the compressed domain. The system has the ability to detect tampering. (Chamlawi et al., 2009), proposed a secure semi-fragile watermarking technique that embed two watermarks in DWT bands, one for authentication and the other for recovery. The scheme is tolerant against JPEG lossy compression and able to detect the tampered areas accurately. All of the previous methods have focused on the authentication of a panchromatic RS image.

In this paper, contrary to the previous research, we propose a novel semi-fragile watermarking method applied to multispectral RS images. The visible bands; Red, Green and Blue (RGB) are exploited to embed the watermark. The three bands are decomposed by four levels integer wavelet transform (IWT). The watermark is a greyscale logo image expanded to 8 bit-planes and XORed with pseudo number sequence to satisfy more security. The bit-planes of watermark are embedded into the bit-planes of the details bands coefficients of the fourth level wavelet transform ( LH4, HL4, HH4 ) for the three visible bands of the RS image.

The rest of the paper is organized as follows: Section 2 reviews the basics of IWT. The semi-fragile watermarking algorithm is discussed in Section 3 and the experiment results under some
attacks on a satellite images are presented and discussed in section 4. Finally, the conclusion is given in Section 5.

2. INTEGER WAVELET TRANSFORM

The first generation of wavelet transform is a floating point transform. If an image with integer pixel values is transformed to wavelet domain and perform some process on the wavelet coefficients (as embedding watermark), the resulting image after inverse transform will not have integer values. Some truncation will be performed to have integer pixel values again as a result the image will not reconstruct exactly. In addition, the computation is complicated and costly. This problem is solved in the second generation of wavelet transform using lifting scheme. Lifting Wavelet Transform (LWT) is an integer to integer mapping proposed by (Sweldens,1998). As this scheme uses lifting instead of convolution, it has many advantages than conventional wavelet transform such as: perfect reconstruction, less memory requirement and speed computation. The lifting scheme consists of three steps (Latha et al., 2008) as shown in Figure 1:

Split step: divides the input data into odd and even elements.
Predict step: predicts the odd elements from the even elements (detail coefficient).
Update step: replaces the even elements with an average (approximate coefficient).

3. METHODOLOGY

3.1 Watermark Generation

For greyscale image, every pixel in the image is represented by 8-bits. Thus, the image can be decomposed into a series of 8-bit planes or 8 binary images. The 1st bit-plane is the least significant bit (LSB) and 8th bit-plane is the most significant bit (MSB). The MSB image contains the most important features of the image and decrease gradually until the LSB image. Our watermark is a greyscale logo image. It is decomposed into 8 bit-planes denoted as \{w0, w1, w2, w3, w4, w5, w6, w7\} as shown in Figure 2. Each binary plane image is XORED with pseudorandom binary sequence PN generated by key K1 to form a new watermark \(\tilde{w}_1:\)

\[
\tilde{w}_1 = w_1 \oplus PN
\]  

where \(\oplus\) is the exclusive OR operation
\(l = 7, 6, ..., 0\) is the bit-plane number of the watermark.

3.2 Watermark Embedding

The whole watermark embedding scheme can be discussed in the following steps as shown in Figure 3:

Step1: let \(I_R, I_G, I_B\) with size MxN are the three visible bands: red, green and blue of the RS image which are used to embed watermark.

Step2: decompose \(I_R, I_G, I_B\) by 4-level IWT into their sub-bands to obtain 9 details sub-bands and one approximate sub-band \(LL_4\) as shown in Figure 4 for each colour band. The \(HL_{4k}^R, LH_{4k}^R\) and \(HH_{4k}^R\) sub-bands are selected to embed watermark where \(k = R, G\) and \(B\). The 4th level details sub-bands are chosen to embed watermark to trade-off between robustness and invisibility.

The watermark embedding operation in these sub-bands is more robust than the lower decomposition level sub-bands and in the same time results in less image degradation than approximation sub-band \(LL_k\). The embedding sub-bands are arranged as following:

\[
HL_{4k}^R, LH_{4k}^R, HH_{4k}^R, HL_{k}^R, LH_{k}^R, HH_{k}^R
\]

Step3: decompose each of the selected sub-bands into 8 bit-planes \(\{b_{8k}^1, b_{8k}^2, ..., b_{8k}^8\}\) where \(b\) is the pit-plane, \(n\) is the sub-band \(HL_{4k}, LH_{4k}\) or \(HH_{4k}\). One of 8 bit-planes of each sub-band is chosen to embed watermark. As adding watermark in the high bit-planes will lead to more degradation in the image and adding watermark in the low bit-planes leads to low robustness, then middle bit-planes are chosen to hide watermark.

Step4: one of bit-plane of the watermark \(\{\tilde{w}_1, \tilde{w}_8, ..., \tilde{w}_9\}\), is added into the \(m^h\) bit-plane of one of the selected sub-bands \(\{b_{mh}^1, b_{mh}^2, ..., b_{mh}^8\}\) respectively according to the following rule:

\[
bw_{mh}^n(i,j) = \begin{cases} 
\text{comp}(b_{mh}^n(i,j)) & \text{if } \tilde{w}_i(i,j) = 1 \text{ and xor } (bw_{mh}^n(i,j), bw_{mh+1}^n(i,j)) = 1 \\
0 & \text{otherwise}
\end{cases}
\]  

where \(bw_{mh}^n(i,j)\) is the watermarked bit at the pixel position \(i,j\) in the \(m^h\) bit-plane of sub-band \(n\) and \(comp =\) the complement of \(b_{mh}^n(i,j)\).
Step 5: once the embedding process is done, the watermarked bit-planes are recomposed into watermarked sub-bands.

Step 6: apply the inverse integer wavelet transform (IIWT) to the modified coefficients to obtain watermarked RS image with color watermarked bands $I_w^R$, $I_w^G$, $I_w^B$.

The key

3.3 Watermark Extraction

Our watermarking technique is blind so it doesn’t require the original host image. The key $k_1$ must be known in the extraction operation. The extraction procedure includes the following steps as shown in Figure 5:

Step1: let $I_{wR}$, $I_{wG}$ and $I_{wB}$ are the three visible watermarked bands of RS image corresponding to red, green, and blue bands, respectively.

Step2: decompose $I_{wR}$, $I_{wG}$ and $I_{wB}$ by 4-level IWT into their sub-bands and choose the $HL_{4,R}$, $HL_{4,G}$, $HL_{4,B}$ and $HH_{4,R}$, $HH_{4,G}$, $HH_{4,B}$ sub-bands used for watermark embedding. The watermarked sub-bands are arranged as following as in the embedding phase: $HL_{4,R}$, $HL_{4,G}$, $HL_{4,B}$, $HH_{4,R}$, $HH_{4,G}$, $HH_{4,B}$

Step 3: decompose each of the selected sub-bands into 8 bit-planes $\{b_{b,k}^n, b_{b,k}^{n+1}, \ldots, b_{b,k}^m\}$ where $b$ is the bit-plane, $n$ is the watermarked sub-band $HL_{4,R}$, $HH_{4,R}$ or $HH_{4,B}$. For each watermarked sub-band, separate the $m^{th}$ bit-plane used in watermark embedding.

Step 4: the watermark bit-planes $\{w'_1, w'_2, \ldots, w'_m\}$ are obtained by using the following rule:

$$w'_l(i,j) = \begin{cases} 1 & \text{if } b_{m,k}^n(i,j) = b_{m+1,k}^n(i,j) \\ 0 & \text{if } b_{m,k}^n(i,j) \neq b_{m+1,k}^n(i,j) \end{cases} \quad (3)$$

where $w'_l(i,j) = \text{the extracted watermark bit at the pixel position } i,j \text{ in the } l^{th} \text{ bit-plane.}$

Step 5: use the secret key $k_1$ to obtain the final watermark $w'_l$ by XORing the extracted watermark bit-planes with $PN$ as in equation (4).

$$w'_l = w'_l \oplus PN \quad (4)$$

Combine the watermark bit-planes to generate the greyscale extracted watermark $w^*$.

3.4 Watermark Authentication and Temper Detection

To analyze the authenticity of the received RS image, the extracted watermark $w^*$ is compared with the original watermark $\tilde{w}$. If they are identical, the RS image is authenticated; otherwise the image is altered intentionally or unintentionally. The normalized correlation ($NC$) is used to measure the similarity between $w^*$ and $\tilde{w}$. The $NC$ is defined as:

$$NC = \frac{\Sigma_i \Sigma_j \tilde{w}^*(i,j)w^*(i,j)}{\sqrt{\Sigma_i \Sigma_j \tilde{w}^*(i,j)^2 \Sigma_i \Sigma_j w^*(i,j)^2}} \quad (5)$$
The $NC$ varies between 0 and 1. According to $NC$ value, we can determine whether the RS image is tampered or not. If $NC=1$, the image is fully authenticated and when the $NC$ value decreases the authenticity decreases. Also, the embedded watermark is used to detect the location of the altered bits.

4. EXPERIMENTAL RESULT

In this section, the performance of the proposed algorithm is estimated. Four color satellite images with different contents and size are selected as test images. The images for ‘Pyramids’ in Egypt and ‘Reliant Stadium’ in Texas are Quickbird satellite images with size 2048x2048, the image ‘Oil Field’ in Iraq is IKONOS satellite images with size 800x768 and ‘Obama’s Inauguration’ in USA is a GeoEye-1 satellite images with size 1024x512. The three color bands of RS images; R, B, and G are watermarked during the test procedures. The watermark shown in Figure 6, is a greyscale image has a size based on the test image.

The Peak-Signal-To-Noise Ratio ($PSNR$) is used to measure the visual quality of watermarked and attacked image (Plaintz and Maeder, 2005) and is defined as:

$$PSNR = 10 \log_{10} \frac{255^2}{\sum_{i=1}^{m} \sum_{j=1}^{n} (I(i,j)-I'(i,j))^2}$$

where $m, n$ = give the size of the image, $I(i,j), I'(i,j) = the$ pixel values at location $(i,j)$ of the original and distorted images respectively.

![Figure 6. Watermark](image)

Table 1. $PSNR$ values of the watermarked image.

<table>
<thead>
<tr>
<th>Image</th>
<th>$PSNR$ (dB)</th>
<th>Band R</th>
<th>Band G</th>
<th>Band B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyramids</td>
<td></td>
<td>40.6993</td>
<td>46.8478</td>
<td>40.6989</td>
</tr>
<tr>
<td>Reliant Stad.</td>
<td></td>
<td>40.6871</td>
<td>46.7659</td>
<td>40.7308</td>
</tr>
<tr>
<td>Oil Field</td>
<td></td>
<td>40.7044</td>
<td>46.8143</td>
<td>40.7613</td>
</tr>
<tr>
<td>Obama’s In.</td>
<td></td>
<td>40.8103</td>
<td>46.9270</td>
<td>40.6803</td>
</tr>
</tbody>
</table>

It is noted that the $PSNR$ for the band G is higher than R and B because the number of bit-planes embedded in band G are less than R and B. For embedding the watermark 8 bit-planes we use three sub-bands from B, three sub-bands from R and two sub-bands from G because the modification on B band is less visible, followed by R band, and finally G band.

The robustness of the proposed algorithm is tested against some unintentional attacks such as some image processing: JPEG compression, Gaussian noise, median filter and sharpen. Table 2 shows the $NC$ results for the previous attacks. From the table we can find that the $NC$ values exceed 0.9 and this is a high value indicates that the extracted watermark is nearly identical to the original watermark. Moreover, to examine our scheme for tamper detection and localization, cut and paste attack is used to change some features of the RS image. As shown in Figure 9, we cut some parts of the images and paste them in other places. The proposed method can detect perfectly the region of the images which are tampered and marked them by black color as shown in Figure 10.

Figure 7 and figure 8 show the original test images and the watermarked images. The $PSNR$ values of the watermarked images are shown in Table 1. These values indicate that there is no visual difference between the original and watermarked image. So our proposed algorithm satisfies the imperceptibility.

5. CONCLUSION AND FUTURE WORK

In this paper, a new semi-fragile watermarking scheme for color satellite images content authentication has been presented. The bit-planes of watermark are embedded into selected bit-planes of details sub-bands of IWT of color bands of RS images. The proposed scheme has examined the integrity of the RS images and this issue is very effective especially in military field. As a result we can create and spread these images.
Figure 7. Tested remote sensing images: (a) Pyramids, Egypt (Quickbird, 2010a). (b) Reliant Stadium, Texas, (Quickbird, 2010b). (c) Oil Field, Iraq (IKONOS, 2010). (d) Obama’s inauguration, USA (GeoEye, 2010).

Figure 8. The watermarked remote sensing images

<table>
<thead>
<tr>
<th>Attack</th>
<th>Pyramids</th>
<th>Reliant Stad.</th>
<th>Oil Field</th>
<th>Obama’s In.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No attack</td>
<td>1</td>
<td>0.9998</td>
<td>0.9997</td>
<td>1</td>
</tr>
<tr>
<td>JPEG compression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality Factor = %90</td>
<td>0.9785</td>
<td>0.9754</td>
<td>0.9821</td>
<td>0.9777</td>
</tr>
<tr>
<td>Quality Factor = %80</td>
<td>0.9681</td>
<td>0.9679</td>
<td>0.9720</td>
<td>0.9675</td>
</tr>
<tr>
<td>Quality Factor = %70</td>
<td>0.9617</td>
<td>0.9638</td>
<td>0.9664</td>
<td>0.9632</td>
</tr>
<tr>
<td>Quality Factor = %60</td>
<td>0.9567</td>
<td>0.9566</td>
<td>0.9625</td>
<td>0.9605</td>
</tr>
<tr>
<td>Gaussian noise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance=.001</td>
<td>0.9870</td>
<td>0.9996</td>
<td>0.9997</td>
<td>0.9997</td>
</tr>
<tr>
<td>Variance=.005</td>
<td>0.9706</td>
<td>0.9963</td>
<td>0.9985</td>
<td>0.9975</td>
</tr>
<tr>
<td>Variance=.01</td>
<td>0.9542</td>
<td>0.9951</td>
<td>0.9957</td>
<td>0.9944</td>
</tr>
<tr>
<td>Sharpen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parameter =.2</td>
<td>0.9835</td>
<td>0.9508</td>
<td>0.9693</td>
<td>0.9632</td>
</tr>
<tr>
<td>Parameter =.4</td>
<td>0.9835</td>
<td>0.9520</td>
<td>0.9689</td>
<td>0.9631</td>
</tr>
<tr>
<td>Parameter =.6</td>
<td>0.9834</td>
<td>0.9525</td>
<td>0.9699</td>
<td>0.9641</td>
</tr>
<tr>
<td>Median filter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3x3</td>
<td>0.9707</td>
<td>0.9635</td>
<td>0.9711</td>
<td>0.9674</td>
</tr>
<tr>
<td>5x5</td>
<td>0.9374</td>
<td>0.9257</td>
<td>0.9375</td>
<td>0.9368</td>
</tr>
</tbody>
</table>

Table 2. The NC results for the different attacks
with more confidence that any alteration can be discovered. From our tests, we can conclude that the proposed method can differentiate between malicious and incidental attacks. Accordingly, the scheme is sensitive to any alteration of the RS images by cut and paste malicious attack. The altered region can be detected and localised perfectly. In addition, the proposed scheme is robust against some common image processing manipulations such as JPEG compression, Gaussian noise adding, sharpen and median filtering.

In the future work, we will try to developing authentication method can recover the tampered region of color RS image and robust to more image processing attacks.

REFERENCES


Devi, P. M., Venkatesan M. and Duraiswamy, K., 2009a. A Fragile Watermarking Scheme for Image Authentication with


