Authentication Algorithm for Color Images using Watermarking Techniques

LUIS ROSALES-ROLDAN, CLARA CRUZ-RAMOS, MARIKO NAKANO-MIYATAKE and HECTOR PEREZ-MEANA
Postgraduate Section, Mechanical Electrical Engineering School, National Polytechnic Institute.
Av. Santa Ana No. 1000, San Francisco Culhuacan, Coyoacán, C.P. 04430.
MEXICO D.F.
rrrosalesr0600@ipn.mx, ccruzra@ipn.mx, mnakano@ipn.mx, hmperezm@ipn.mx.

Abstract: - In this paper we present a watermarking algorithm for color image content authentication with localization and recovery capability of the modified areas. We use a halftone image generated by the Floyd-Steinberg kernel as an approximate version for the luminance channel of the host image and an encoded version of the color information (chrominance channels). We adopt this halftone image and the encoded chrominance channels as watermark sequences and embed it using the quantization watermarking method into the sub-band LL, HL and LH of the Integer Wavelet Transform (IWT) of the host image. Due to the fact that the watermark is embedded into these sub-bands of IWT, the proposed method is robust to JPEG compression. Moreover, we employ a Multilayer Perceptron Neural Network (MLP) as inverse halftoning process to improve the recovered image quality. Using the extracted luminance channel, the shape of the modified area is estimated by the MLP. The experimental results demonstrate the effectiveness of the proposed scheme.

Key-Words: - Image Authentication, Color Image, Halftone, IWT, Recovery Capability, Tamper Detection, NCD, Neural Network.

1 Introduction

During the last few years, we have seen a tremendous growth in technological advances creating new and better tools in areas such as informatics, electronic and telecommunications. These advances have a strong impact on the people’s life, for example it is quite common to take pictures everywhere and every time using his/her cell phones with digital cameras. And also 3125 pictures per minute are uploaded to the cloud to be share in any social network. However these digital pictures can be easily modified using computational drawing tools, such as Photoshop or GIMP without causing any distortion. Nowadays almost all digital images are color images and until now these color images have been treated as an extension of gray scale images, however due to that the vulnerability and importance between color and luminance information are very different, the color information must be treated in a different manner.

Watermarking algorithms are generally used for content origin identification, copy protection, illegal copy tracking, fingerprinting and content access control, which are based either on an additive, multiplicative or a quantization process. Although several methods have been proposed to watermark grey level images, only a few methods have been designed for color images. According to [1] the color information can be add using histogram, frequency domain transform or spatial domain processing. The most common is watermarking color image through a transform domain (DCT, DFT or DWT), witch embeds the marking information into the coefficients of the transform.

Some transform watermarking schemes based on the Discrete Wavelet Transform (DWT) have been proposed [2-6]. The main advantage of these schemes is that it takes into account the image characteristics, witch make it possible to embed more strongly a watermark sequence. The authors in [6] control the imperceptibility and the robustness of watermarked images using the contextual entropies of the host wavelet coefficients. The scheme proposed by [7] use a non-linear Supported Vector Machine (SVM) to exploits the color statistics of a first-order and high-order wavelet statistics, simplifying the detection of the mark. However this approach is applied independently to each color component.

In this paper we present a color-watermarking algorithm for image content authentication with localization and recovery capability of the modified areas. We use a halftone image generated by the Floyd-Steinberg kernel as an approximate version for the luminance channel of the host image and an encoded version of the color information. We adopt this information as
watermark sequences and embed it using the quantization watermarking method into the sub-band LL, HL and LH of the Integer Wavelet Transform (IWT) of the host image. Due to the fact that the watermark is embedded into these sub-bands of IWT, the proposed method is robust to JPEG compression. Moreover, we employ a Multilayer Perceptron neural network (MLP) as inverse halftoning process to improve the recovered image quality [8]. Using the extracted luminance channel, the shape of the modified area is estimated by the MLP. The experimental results demonstrate the effectiveness of the proposed scheme.

The rest of this paper is organized as follows. Section 2 describes the proposed algorithm and experimental results are presented in Section 3. Finally Section 4 concludes this work.

2 Proposed Watermarking-based Algorithms

The proposed authentication algorithm is composed by three stages: self-embedding, authentication and recovery stage as shown in Fig.1.

![Fig.1 General scheme of the proposed algorithm.](image)

2.1 Self-embedding Stage

The self-embedding stage has two different processes, the first one is the watermark sequence generation and the second one is the embedding process.

In the watermark sequence generation the original color image is converted from RGB into YCbCr color space, then the luminance channel \( Y \) is down-sampled with half size in height and width to obtain two new chrominance channels \( (Cb^2 \) and \( Cr^2 \)), these chrominance are segmented into 8x8 pixels blocks, which are transformed by using the 2D-IDCT with a quantization matrix with compression factor \( Q=50 \). Form each of these blocks; the DC and the first two AC values are coding using 7 bits and 8 bits, respectively. Finally a parity bit is added to create a sequence of 16 bits for each 8x8 pixels block. Creating two color watermarks, \( BinQCb \) and \( BinQCr \), respectively. Transformation Process

In the embedding process, the color original image is converted from RGB to YCbCr color space, then the luminance channel \( Y \) is decomposed using the IWT to obtain four sub-bands: LL, LH, HL and HH. The \( HTY \) watermark sequence is embedded into the sub-band LL using the quantization watermarking method. On the other hand the sub-band LH and HL are decomposed using the IWT twice to obtain LL2 and LL4 respectively. The \( BinQCb \) and \( BinQCr \) watermark sequences are embedded into the sub-band LL2 and LL4 using the same quantization watermarking method, respectively. The quantization embedding formula used for the algorithm is given by:

\[
  w_k = \begin{cases} 
  c_{ij} = v_i & \text{if } \left| c_{ij} - v_i \right| \leq \left| c_{ij} - v_2 \right| \\
  c_{ij} = v_2 & \text{otherwise}
  \end{cases}
\]

(1)

where

\[
  v_i = \begin{cases} 
  \frac{|c_{ij}|}{2S} \times 2S & w_k = 0 \\
  \frac{|c_{ij}|}{2S} \times 2S + S & w_k = 1
  \end{cases}
\]

(2)

and \( w_k \) is the \( k \)-th watermark bit, \( c_{ij} \) and \( c_{ij}' \) are the original and the watermarked IWT coefficients, respectively, and \( S \) is the quantization step size. Finally we obtained the watermarked image applying inverse IWT to the watermarked sub-bands \( LL2, LL4 \) and \( LL \). This stage is shown in Figure 3.
The simplest way to realize gray-scale image conversion is to use a Gaussian low-pass filter given by (4). To recover the color information embedded into the suspicious color image the watermark sequences extracted from the LL2 and LL4 sub-bands are decoded using the inverse 2D-IDCT and the matrix quantization $Q$ for each 8x8 block. Creating the suspicious preprocessed image (SPI).

$$F_{2} = \frac{1}{11.566} \begin{bmatrix} 0.1628 & 0.3215 & 0.4035 & 0.3215 & 0.1628 \\ 0.3215 & 0.6352 & 0.7970 & 0.6352 & 0.3215 \\ 0.4035 & 0.7970 & 1 & 0.7970 & 0.4035 \\ 0.3215 & 0.6352 & 0.7970 & 0.6352 & 0.3215 \\ 0.1628 & 0.3215 & 0.4035 & 0.3215 & 0.1628 \end{bmatrix}$$

Next we generate a halftone image from the luminance channel ($Y$) of the suspicious watermarked color image converted and it is re-converted in a gray-scale image using the same Gaussian low-pass filter. This inverse halftoning is the simplest method, even though it produces low quality gray-scale image. The other two channels ($Cb$ and $Cr$) form the suspicious watermarked color image converted are encoded as the watermark generation, then, their are decoded using the inverse 2D-IDCT and the matrix quantization $Q$ for each 8x8 block. Creating the suspicious preprocessed image (SPI).

In this stage, an accurate detection of the modified areas is important; therefore high quality of the gray-scale image is not necessary. Then both images (EWI and SPI) are compared each other to localize the modified areas. To do this we employed a block-wise strategy, in which the comparison is carried out in each block of NxN pixels and the normalized color difference (NCD) of each block is calculated by (5) and it is compared with a predetermined threshold value $Th$.

$$D_{k} = NCD(EWI, SPI)k = 1, 2, 3, ..., TB$$

where $TB$ is the total of block of NxN pixels.

2.3 Recovery Stage

If the authentication stage shows that some blocks of the suspicious color image are tampered, then the recovery stage will be triggered. In this stage we will use as input data, the down-sampled suspicious watermarked image, its halftone version, the information about modified blocks and the extracted halftone image (signal HT in Fig.4). Also the other two extracted watermarks from de sub-band HL and LH. (signals Cb1 and Cr1 in Fig.4). According to [8], in this stage we firstly use the down-sampled luminance channel ($Y$) of suspicious image and its halftone version to train MLP by the Backpropagation (BP) algorithm. This recovery stage is shown in Fig.5 and the MLP used to estimate the gray-scale image is shown in Fig.6.
3 Experimental Results

In this section the performance of the proposed algorithm, is evaluated from several points of view, such as the watermark imperceptibility, watermark robustness against JPEG compression, tamper detection accuracy and recovery capability. The watermark imperceptibility and robustness of the proposed algorithm are strongly depends on step size value $S$ used in watermark embedding and extraction algorithms given by (2) y (3), respectively. After an analysis of image quality and watermark robustness was determined the value of the Step-Size, which is 7.

3.1 Watermark Imperceptibility

The watermark imperceptibility of the proposed algorithm is evaluated using 25 color images. Table 1 shows the Peak Signal to Noise Ratio (PSNR) and the Normalized Color Difference (NCD) of the 25 watermarked color images respect to their original versions.

<table>
<thead>
<tr>
<th>Paso de Cuantización</th>
<th>PSNR</th>
<th>NCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>37.5280</td>
<td>0.0227</td>
</tr>
<tr>
<td>6</td>
<td>36.0530</td>
<td>0.0287</td>
</tr>
<tr>
<td>7</td>
<td>34.8731</td>
<td>0.0308</td>
</tr>
<tr>
<td>8</td>
<td>33.7995</td>
<td>0.0347</td>
</tr>
<tr>
<td>9</td>
<td>32.7661</td>
<td>0.0396</td>
</tr>
<tr>
<td>10</td>
<td>31.8314</td>
<td>0.0494</td>
</tr>
</tbody>
</table>

Fig.7 Image quality comparison. (a) Original color image, (b) color image by Gaussian low-pass filter (19.62dB), (c) color image by MLP (20.75dB).

Fig.8 Watermark imperceptibility. (a, b) Original color images, (c, d) Watermarked color images.
Two original color images and their watermarked color images generated by the proposed algorithms are given in Fig.7. From the Fig.8, the PSNR and the NCD values in the table, which is approximately 35 dB and 0.03, respectively; we can conclude that the proposed algorithm provides high imperceptibility of the watermark sequences. PSNR and NCD values of the watermarked color images 35.04 and 35.15 dB, 0.0262 and 0.0241, respectively.

3.2 Watermark Robustness

The main attack that not modified the content of a color image is the JPEG compression. JPEG is the most common image format used by digital cameras and other image capture devices, reducing the size of the image files.

The watermark robustness against JPEG compression of the proposed algorithm is evaluated varying the JPEG quality factor. Table 2 shows the relationship between the PSNR and NCD of the extracted watermark color image after the JPEG compression and the marked one.

Table 2. Watermark robustness against JPEG compression.

<table>
<thead>
<tr>
<th>Step Size</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9.8872</td>
<td>14.4350</td>
<td>22.4565</td>
</tr>
<tr>
<td>6</td>
<td>11.1746</td>
<td>16.0178</td>
<td>22.3495</td>
</tr>
<tr>
<td>7</td>
<td>12.4100</td>
<td>19.0537</td>
<td>22.3305</td>
</tr>
<tr>
<td>8</td>
<td>13.0906</td>
<td>19.7280</td>
<td>22.2543</td>
</tr>
<tr>
<td>9</td>
<td>14.6977</td>
<td>21.3703</td>
<td>22.1448</td>
</tr>
<tr>
<td>10</td>
<td>15.5182</td>
<td>21.6217</td>
<td>22.0306</td>
</tr>
<tr>
<td>NCD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.9324</td>
<td>0.5364</td>
<td>0.1444</td>
</tr>
<tr>
<td>6</td>
<td>0.7802</td>
<td>0.4431</td>
<td>0.1473</td>
</tr>
<tr>
<td>7</td>
<td>0.6941</td>
<td>0.2871</td>
<td>0.1461</td>
</tr>
<tr>
<td>8</td>
<td>0.6292</td>
<td>0.2537</td>
<td>0.1470</td>
</tr>
<tr>
<td>9</td>
<td>0.5304</td>
<td>0.1815</td>
<td>0.1482</td>
</tr>
<tr>
<td>10</td>
<td>0.4758</td>
<td>0.1708</td>
<td>0.1523</td>
</tr>
</tbody>
</table>

According to the previous tests, compression with quality factor 100% is enough for a color image, because the generated color image is 75% smaller than the original one. The Fig.9 shows an example of the watermark color image extracted after JPEG compression with quality factor 100%, PSNR 23.72 dB and NCD 0.1086.

Fig.9 Watermark robustness. (a) Watermarked color image, (b) watermarked color image extracted after JPEG compression 100%.

3.3 Tamper Detection and Recovery Capability

In the proposed algorithm, the tamper detection is performed by (5). The NCD value $D_k$ is calculated in each block with $N \times N$ pixels and then is compared with a predefined threshold value $Th$ to determine if the block is tampered or not. The block size $N \times N$ is set to $8 \times 8$, and the threshold value $Th$ is determined taking account of the false alarm and false negative error rates.

Fig.10 (a) Watermarked color image, (b) tampered version (3.81%), tampered image with the detected tampered blocks is shown by (c), and (d) show the recovered image (25.11 dB).

Although using an adequate threshold value $Th$ determined as mentioned above, many isolated blocks are detected as tampered one, however an isolated block with size $8 \times 8$ pixels is visually insignificant, and then using the connected
component labeling algorithm [10], the isolated blocks are eliminated, however if false alarm error is present, the recovered area will be the same with lower quality. An example of the tamper detection and recovery process of the proposed algorithm is shown in the Fig.10.

4 Conclusions

In this paper we proposed a color watermarking-based algorithm for tamper detection and recovery, where, the watermark embedding is carried out in the Integer Wavelet Transform domain. In this algorithm, a halftone image and a coded version of the color information are used for tamper detection and recovery of the tampered region, which ones are embedded into the image as watermark sequences using qualification watermarking algorithm. The proposed algorithm was evaluated from the watermark imperceptibility and robustness. The average PSNR and NCD of several watermarked color images respect to their original versions using an adequate step size value indicates that the embedded watermark is imperceptible by Human Visual System. Also simulation results showed that the embedded watermark is robust to JPEG compression with a quality factor larger than 80%. The use of the MLP trained by BP algorithm increases the quality of the recovered image and the simulation results showed that the proposed method can detect and recover correctly the modified areas.

References