A Semi-Fragile Watermarking Scheme for Color Image Authentication

M. Hamad Hassan and S.A.M. Gilani

Abstract—In this paper, a semi-fragile watermarking scheme is proposed for color image authentication. In this particular scheme, the color image is first transformed from RGB to YST color space, suitable for watermarking the color media. Each channel is divided into 4×4 non-overlapping blocks and its each 2×2 sub-block is selected. The embedding space is created by setting the two LSBs of selected sub-block to zero, which will hold the authentication and recovery information. For verification of work authentication and parity bits denoted by 'a' & 'p' are computed for each 2×2 sub-block. For recovery, intensity mean of each 2×2 sub-block is computed and encoded upto six to eight bits depending upon the channel selection. The size of sub-block is important for correct localization and fast computation. For watermark distribution 2D-Torus Automorphism is implemented using a private key to have a secure mapping of blocks. The perceptibility of watermarked image is quite reasonable both subjectively and objectively. Our scheme is oblivious, correctly localizes the tampering and able to recovery the original work with probability of near one.

Keywords—Image Authentication, YST Color Space, Intensity Mean, LSBs, PSNR.

I. INTRODUCTION

In past decade, there has been exponential growth in the use of digital multimedia contents. The wideband networks made the exchange of multimedia contents, easy and fast. On the other hand, the availability of powerful image processing tools made it easy for user to do even imperceptible changes in the original work. As a result image authenticity becomes greatly threatened. Generally image authentication verifies the integrity of a digital image. Several questions may be asked about the authenticity of a work, e.g. has the work been altered in any way whatsoever, has the work been significantly altered, what parts of work have been altered, can an altered work be recovered. Although other techniques for answering these questions exist, Digital Watermarking Techniques may be useful because they do not require auxiliary data and they undergo the same transformations as the cover work.

Depending on the application, digital watermarking techniques can be classified into two main categories; Robust and Fragile Watermarking Techniques. The former is mainly used for copy right protection and fingerprinting applications, in which the goal of watermark is to sustain under all kinds of attacks that intend to remove the watermark while preserving the perceptual quality of the original media. The latter is used for data authentication and is sensitive to any kind of processing that may occur. A fragile watermark is very sensitive and is designed to detect every possible change in a marked image, but in most multimedia applications, minor data modifications are acceptable as long as the content is authentic. A semi-fragile watermark is robust to acceptable content preserving manipulations such as lossy compression while fragile to malicious distortions such as content modification.

In this paper, we propose a semi-fragile watermarking scheme for color image authentication. The given color image is first transformed from RGB to YST color space. This new color space is exclusively designed by Francesco et al. [3] for watermarking based applications. Details of this color space are discussed in Section III of this paper. The Y channel corresponds to the luminance component while S & T channels correspond to the chrominance component of color image. Since YS ⊥ T therefore T channel is selected for embedding the authentication bits apart from the recovery bits, whereas, YS channels hold, only the recovery information. Each channel is divided into 4×4 non-overlapping blocks and its each 2×2 sub-block is selected. The embedding space for watermark is created by setting the LSBs of selected sub-block to zero. The LSBs plane is required to hold the corresponding blocks authentication and recovery information. For recovery, the intensity mean of each 2×2 sub-block is computed and encoded upto six bits in case of T channel and upto eight bits in case of YS channels. The authentication and parity bits i.e. ‘a’ & ‘p’ are computed for each sub-block using equation (3) and (4) as discussed in Section III of this paper. The size of sub-block is important for correct localization and fast computation. The corresponding block that will hold the source block information is determined by implementing 2D-Torus Automorphism presented by G. Voyatzis et al. [4], using a private key to have secure mapping of blocks by using equation (2). The perceptibility of watermarked image is quite high as the PSNR value for all the tested images is greater than 38 dB. Our scheme is oblivious, correctly localizes the tampering with full recovery of the original work.

The rest of the paper is organized as: Section II briefly states the related work, Section III explains about the
necessary background. Section IV explains the proposed scheme, Section V demonstrates the simulation results, and Section VI derives the concluding remarks.

II. RELATED WORK

The earliest scheme for image authentication was proposed by Walton [5] that uses a key dependent pseudo-random walk on the image. The check-sum is obtained by summing the numbers determined by the seven most significant bits (MSBs) and taking a remainder operation with a large integer \( N \). The computed check-sum is then inserted in a binary form in the LSBs of the selected pixels. The method is very fast and on average modifies only half of the pixels by one gray level. The check-sum approach provides a very high probability of tamper detection, but cannot distinguish between an innocent adjustment of brightness and replacing a person’s face.

Van Schyndel et al. [6] modify the LSBs of the pixels by adding extended m-sequences to rows of pixels. The sequences are generated with a linear feedback shift register. For an \( N \times N \) image, a sequence of length \( N \) is randomly shifted and added to the image rows. The phase of the sequence carries the watermark information. A simple cross-correlation is used to test for the presence of the watermark.

Wolfgang and Delp [7] extended van Schyndel’s work and improved the localization properties and robustness.

Fridrich [8]-[10] proposed schemes where in which image is divided into 8\( \times 8 \) blocks and each block is DCT (Discrete Cosine Transform) transformed. Specified number of the lowest frequency DCT coefficients are quantized using a given quantization table \( Q \) that corresponds to the 50% of JPEG quality. The encoded bits were then embedded into LSBs for authentication & recovery.

Detailed survey on authentication schemes proposed by research community can be found in Liu, Z.D Que et al; in [11].

Phen et al [1] & Jagdish et al [2] proposes hierarchical based scheme in which intensity mean of blocks are embedded into randomly chosen block LSBs for authentication and recovery but for gray scale image authentication.

In this paper we are presenting scheme for color image authentication using state of the art color space i.e. YST presented by Francesco et al. [3] and exclusively designed for watermarking based applications. In RGB color space, channels are highly correlated but in YST color space they are orthogonal to each other. Given color image is first transformed from RGB to YST color space using set of linear transformation matrix given by equation (1). Geometrically \( YS \perp T \) therefore \( T \) channel is selected to hold authentication information apart from recovery information. Rest of channels is deployed to hold only the recovery information. Each channel is divided into 4\( \times 4 \) non-overlapping blocks and its 2\( \times 2 \) sub-block is selected to set the two LSBs to zero. The LSBs plane is required to hold the corresponding block’s authentication & recovery information. For recovery, the intensity mean of each 2\( \times 2 \) sub-block is computed and encoded upto six bits in case of \( T \) channel and upto eight bits in case of \( YS \) channels. The authentication and parity bits i.e. ‘\( a \)’ & ‘\( p \)’ are computed for each 4\( \times 4 \) block’s sub-block using equation (3) and (4) respectively as discussed in Section III of this paper. The size of sub-block is important for correct localization and fast computation. The target block that will hold the source block information is determined by implementing the 2D-Torus Automorphism presented by G. Voyatzis et al. [4], using private key to have secure mapping of blocks by using the equation (2). The perceptibility of watermarked image is quite high as the PSNR value for all the tested images is greater than 38 dB. Our scheme is oblivious, correctly localizes the tampering in a work with full recovery of the original work.

III. BACKGROUND KNOWLEDGE

YST Color Space: The selection of color space is very important step in watermarking based applications. In this regard we considered YST color space, exclusively designed and recommended by Francesco et al. [3]. A color space is notation by which we specify colors i.e. human perception of the visible electromagnetic spectrum.

The RGB color space is good for image display but is not the best choice when analyzing images using the computer. The main disadvantage of the RGB color space is high correlation between its components. The value of cross-correlation between the \( B \) and \( R \) channel is numerically about 0.78, 0.98 between the \( R \) and \( G \) channel and 0.94 between the \( G \) and \( B \) channels respectively. Because of this high correlation between channels, the RGB domain is not suitable for image processing techniques, especially for watermarking the color media. The potential of these three channels can be exploited for the applications of watermarking, by decreasing the correlation among them.

Other colors spaces too exist which possess this feature of separating the luminance component from the chrominance component, such color spaces includes YCbCr, YUV etc. where \( Y \) corresponds to the brightness portion of an image while \( Cr \), \( Ch \) and \( U, V \) corresponds to the chrominance (color) components of the color image.

In case of YST color space, \( Y \) corresponds to the brightness component as before while \( S \) and \( T \) channels correspond to the chrominance component of the color image. The new color space satisfies all the principal conditions that are:

i) The brightness must be the same to that of other two color spaces i.e. \( YUV \) and \( YCbCr \).

ii) One component i.e. \( S \) must be ad hoc created to match the vector corresponding to the skin color.

iii) The transformation should be reversible.

The two components \( Y \) and \( S \) form an angle of 52° because they were generated without imposing any orthogonal criterion. The \( T \) component is identified by the Gram-Schmidt procedure in order to have a component that is orthogonal to the plane containing \( Y \) and \( S \) component, hence \( YS \perp T \). Thus we have set of linear transformation matrix to convert color image from \( RGB \) to \( YST \) color space given by equation (1).
The diagrammatic representation of \( YST \) color space is given by:

\[
\begin{pmatrix}
Y \\
S \\
T
\end{pmatrix} =
\begin{pmatrix}
0.299 & 0.587 & 0.114 \\
-0.147 & -0.289 & 0.436 \\
0.615 & -0.515 & 0.100
\end{pmatrix}
\begin{pmatrix}
R \\
G \\
B
\end{pmatrix}
\]

\( (1) \)

\textbf{2D-Torus Automorphism:} 2D-Torus Automorphism can be considered as a permutation function or spatial transformation of a plane region. This transformation can be performed using the \( 2 \times 2 \) matrix \( A \) with constant elements. A point \((x, y)\) can be transformed to new point \((x', y')\) using equation (2).

\[
\begin{pmatrix}
x' \\
y'
\end{pmatrix} = \begin{pmatrix}
1 & 1 \\
k & k + 1
\end{pmatrix} \begin{pmatrix}
x \\
y
\end{pmatrix} \mod N
\]

Where \((x, y), (x', y') \in [0, N-1] \times [0, N-1], N\), the number of blocks in each dimensions, and \( k \in [0, N-1] \) is a private key. For in depth understanding of how 2D - Torus Automorphism works, reader is recommended to follow paper presented by G. Voyatzis et al. [4].

\textbf{IV. PROPOSED SCHEME}

\textbf{Pre-Processing of Image:}

Let \( C \) be the color image in \( RGB \) color space with size \( MxN \). This color image is first transformed from \( RGB \) to \( YST \) color space using equation (1).

\textbf{Watermark Generation}

1. Select each channel and divide it into non-overlapping \( 4 \times 4 \) blocks and set the two \( LSBs \) of its each \( 2 \times 2 \) sub-block to zero.
2. Compute the intensity mean of each \( 2 \times 2 \) sub-block and encode it up to six binary bits in case of \( T \) channel but encode it up to eight binary bits for \( YS \) channels.
3. Select the \( T \) channel and calculate the authentication and parity bits; ‘\( a \)’ and ‘\( p \)’ bits as follows:

\textit{Calculating Authentication bit}

\[
a = \begin{cases} 
1, & \text{if } sB_\text{mean} > B_\text{mean} \\
0, & \text{otherwise}
\end{cases}
\]

\( (3) \)

\textit{Where:} \( B \): intensity mean of \( 4\times4 \) block \( sB \): intensity mean of \( 2\times2 \) sub-block

\textit{Calculating Parity bit}

\[
p = \begin{cases} 
1, & \text{if } \# \text{ of } 1's \text{ is odd, in six bit code of } sB_\text{mean} \\
0, & \text{otherwise}
\end{cases}
\]

\( (4) \)

\textbf{Watermark Embedding}

1. Select the \( T \) channel of image \( M \), divide it into non-overlapping \( 4\times4 \) blocks & perform a secure mapping of blocks using equation (2) with a private key.
2. Embed the watermark into each mapped \( 4\times4 \) block’s, \( 2\times2 \) sub-block two \( LSBs \) of \( T \) channel as:

\[
\begin{pmatrix}
\text{Pixel 1} & \text{Pixel 2} & \text{Pixel 3} & \text{Pixel 4} \\
\text{LSB#2} & \text{LSB#1} & i2 & i2 \\
i3 & i4 & i5 & i6
\end{pmatrix}
\]

\textbf{Fig. 2 T Channel’s 2x2 sub-block LSBs Embedding}

3. Embed the watermark in each mapped \( 4\times4 \) block’s, \( 2\times2 \) sub-block two \( LSBs \) of \( YS \) channels as:

\[
\begin{pmatrix}
\text{Pixel 1} & \text{Pixel 2} & \text{Pixel 3} & \text{Pixel 4} \\
\text{LSB#2} & \text{LSB#1} & i1 & i2 \\
i3 & i4 & i5 & i6 \\
i7 & i8
\end{pmatrix}
\]

\textbf{Fig. 3 YS Channel’s 2x2 sub-block LSBs Embedding}

4. Once the embedding process is done, concatenate all the channels and transform the image from \( YST \) domain to \( RGB \) by taking inverse transform of equation (1) to have the watermarked image.

\textbf{Watermark Extraction & Tamper Detection}

1. Let \( W \) be the watermarked image in \( RGB \) color space with size \( MxN \). This color image is first transformed from \( RGB \) to \( YST \) color space using equation (1).
2. Select \( T \) channel, divide it into non-overlapping \( 4\times4 \) blocks and select its \( 2\times2 \) sub-blocks to extract the
authentication and parity bits from the designated LSBs as demonstrated in Fig. 2.

3. After extraction of desired bits from T channel, set the two LSBs to zero.

4. Compute the intensity mean of each 2×2 sub-block and encode it up to six binary bits.

5. Calculate the authentication and parity bits; ‘a’ & ‘p’ bits using equation (3) & (4) respectively.

6. Compare the extracted & generated ‘a’ & ‘p’ bits; if they are found same, the tested image is authentic otherwise tampered.

7. In case of forged image, identify those blocks that are tampered by setting their pixels value to zero.

**Recovery of Tampered Work**

1. In case of T channel, simply assign the earlier computed intensity mean in previous phase to the tampered block each pixel.

2. In case of YS channels, divide each channel into 4×4 non-overlapping blocks and identify the corresponding source block for the tampered block using equation (2) with the same private key as used in embedding phase.

3. Extract the recovery bits from each corresponding 2×2 sub-block LSBs, for each 4×4 block, as demonstrated in Fig. 3 and assign it to each pixel of the tampered block.

V. RESULTS

The simulations were conducted on Intel machine with 2.4 GHz processor and 512 MB of RAM. MATLAB 7.0 & Photoshop 7.0 was used for implementation of proposed scheme and image processing operations respectively.

**PSNR Measurement:**

One commonly used measure to evaluate the imperceptibility of the watermarked image is the peak signal to noise ratio (PSNR) which is given by equation (5).

\[
PSNR = 10 \cdot \log_{10} \left( \frac{255}{\text{MSE}} \right) (dB)
\]  

**TABLE I:** QUALITY MATRIX (PSNR)

<table>
<thead>
<tr>
<th>Image</th>
<th>Format</th>
<th>Size</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>tiff</td>
<td>200x200</td>
<td>44.36</td>
</tr>
<tr>
<td>Watch</td>
<td>tiff</td>
<td>200x200</td>
<td>44.34</td>
</tr>
<tr>
<td>F16</td>
<td>tiff</td>
<td>200x200</td>
<td>44.39</td>
</tr>
<tr>
<td>Baboon</td>
<td>tiff</td>
<td>200x200</td>
<td>44.13</td>
</tr>
<tr>
<td>Opera</td>
<td>tiff</td>
<td>256x256</td>
<td>44.17</td>
</tr>
<tr>
<td>Waterfall</td>
<td>tiff</td>
<td>256x256</td>
<td>44.12</td>
</tr>
</tbody>
</table>

In Fig. 4, caption (a), (b), (c), (d) and (e) shows the original image, watermarked image, tampered image, detected image and recovered image respectively.
VI. CONCLUDING REMARKS

In this paper, we proposed a semi-fragile watermarking scheme for color image authentication. The given color image is first transformed from RGB to YST color space suitable for watermarking based applications. The Y channel corresponds to the luminance component while S, T corresponds to chrominance components of color image. As $\text{YST} \perp$, therefore T channel is selected for embedding the authentication bits apart from recovery bits, whereas, YS channels hold, only the recovery information. Each channel is divided into $4 \times 4$ non-overlapping blocks and its $2 \times 2$ sub-block is selected. The LSBS plane is created by setting the value of selected LSBs to zero. The LSBS plane is required to hold the corresponding blocks authentication and recovery information. Authentication and parity bits i.e. ‘a’ and ‘p’ are computed for each sub-block. For recovery, the intensity mean of each $2 \times 2$ sub-block is computed and encoded upto six bits in case of T channel and upto eight bits in case of YS channels. The size of sub-block is important for correct localization and fast computation. The target block that will hold the source block information is determined by implementing 2D-Torus Automorphism, using a private key to have secure mapping of blocks. The perceptibility of watermarked image is quite high as the PSNR value for all the tested images is greater than 38 dB. Our scheme is oblivious, correctly localizes the tampering with full recovery of the original work.  In future we will investigate the proposed scheme for VQ counterfeit attacks and other illegitimate attacks to check the degree of survival of this scheme.

REFERENCES