Semi-fragile watermarking for HGI image compression

Alina Bavrina
Samara National Research University; Image Processing Systems Institute of RAS - Branch of the FSRC
"Crystallography and Photonics" RAS
Samara, Russia
bavrina@mail.ru

Victor Fedoseev
Samara National Research University; Image Processing Systems Institute of RAS - Branch of the FSRC
"Crystallography and Photonics" RAS
Samara, Russia
vicanfed@gmail.com

Abstract—A novel semi-fragile watermarking system adopted for the HGI image compression algorithm is proposed. The watermarking method exploits the hierarchical structure of the image when embedding and replaces post-interpolation residual quantization inside HGI compression with a special quantizer based on quantization index modulation. As a result, the protected image became robust to HGI compression with a tunable quality parameter. Several experiments have shown the ability of the proposed watermarking system to protect images with high quality in terms of PSNR. We also investigate the accuracy of local distortion detection. As a result, a trade-off between image quality and forgery detection accuracy has been found.

Keywords—digital image processing, digital watermarks, image compression, hierarchical grid interpolation method

I. INTRODUCTION

Nowadays, the problem of image (and more specifically, remote sensing image) protection against malicious distortions plays an important role.

Satellite and drone images are increasingly used in various fields of industry, agriculture, in the prevention of natural disasters, in the military sphere, and in the media [1]. Modern image processing tools, which include not only raster editors but also artificial intelligence tools such as generative adversarial neural networks, allow users to create fake images or their fragments that are practically indistinguishable from real ones [2, 3]. Distribution of such fake images can have serious economic and political consequences [4, 5].

One way to protect an image from tampering is to embed a fragile or semi-fragile digital watermark [6, 7]. Fragile watermarks are destroyed after any modifications of protected data. Therefore, if there is a set of allowable modifications (such as compression, or cropping, or color correction, etc.), it is better to use semi-fragile watermarks. They are resistant to the allowable transformations and are destroyed by any others [8]. The difference between the embedded watermarks and those extracted at the authentication stage could evidence illegal changes in the protected image.

Remote sensing images usually have high spatial and/or spectral resolution. Therefore they are usually stored and transmitted in a compressed form. Consequently, the “allowable” transformations often include distortions arising from compression. A watermarking system (we use this term to determine a set of algorithms for watermark embedding and extraction [8]) designed for compressed image protection must be resistant to distortions caused by a corresponding compression algorithm. That is, the embedded watermark should be extracted with high accuracy from compressed data.

In this paper, we consider a hierarchical grid interpolation (HGI) compression method, which shows high performance for still images and especially for remote sensing data [9, 10]. This method has two main advantages: the ability to control the compression error and the ability of hierarchical access to data. These properties make the method attractive for applications in areas where the accuracy of image restoration after compression is important, for example, in remote sensing or when processing medical images.

There are some examples of semi-fragile watermarking systems adopted for different compression formats. More than two dozens were developed for JPEG ([11, 12], many algorithms are compared in the review paper [7]). We can also mention papers [13, 14, 15]. Paper [15] contains an overview of existing watermarking systems for the H.264 video. However, we did not find any example of a watermarking system adopted for HGI. This fact could be explained by a limited HGI usage by the academic community. However, both its importance in practice for remote sensing data storage and its closeness to some other hierarchical compression methods based on wavelets or quadtrees (such as [17-21]) make it actual the study of HGI watermarking.

The paper proposes a semi-fragile watermarking system based on the QIM (Quantization Index Modulation) technique [22, 23], consistent with the HGI compression algorithm. The idea is to use the hierarchical structure of the image when embedding and to replace an HGI quantizer with a QIM-based quantizer.

The parameters of the proposed system make it possible to find the compromise between the watermarking distortions and the robustness to certain attacks (the paper considers two attacks of image compression and local alteration).

The rest of the paper is organized as follows. Section 2 describes the HGI compression method, necessary for a better understanding of the proposed watermarking system, which is described in Section 3. The next section contains some numerical experiments to study the characteristics of the proposed system. Finally, conclusion and future work follow in Section 5.

II. THE HGI COMPRESSION METHOD

The HGI compression method is based on the representation of an image $I(m, n)$ as a union of hierarchical levels:
\[
I(m, n) = \bigcup_{l=1}^{L} I_l(m, n),
\]
where \( I_{l-1}(m, n) \) are the highest level samples taken at a distance of \( 2^{l-1} \) at both coordinates. The following equation specifies the lower levels:
\[
I_l = \{ I_l(m, n) \} \setminus \{ I_{l-1}(m, n) \}.
\]

During compression, at first, samples of the current level are interpolated by higher-level samples. The following stages of the compression procedure are quantization of the interpolation errors, reconstruction of samples (for use at lower levels), statistical coding of post-interpolation residuals, and their storage in an archive. Samples of the highest level are stored in the archive unchanged.

Let us consider these stages in more detail. Let \( l \) be a current level.

A. Interpolation

Interpolation of samples at level \( l \) is based on samples of higher levels that have already passed the quantization and reconstruction procedure:
\[
\hat{I}_l(m, n) = P_{\text{GAI}} \left( \bigcup_{k=1}^{L-1} \{ \hat{I}_k(m, n) \} \right),
\]
where \( P_{\text{GAI}}(...) \) is an interpolation function.

B. Calculation of post-interpolation residuals

Calculation of the differences between true sample values and those obtained by the interpolation:
\[
R(m, n) = I_l(m, n) - \hat{I}_l(m, n).
\]

C. Quantization of post-interpolation residuals

\[
Q_{\text{GAI}}(R(m, n)) = Q_{\text{GAI}}(R, \epsilon_{\text{GAI}}),
\]
where \( Q_{\text{GAI}} \) is a quantizer that guarantees the preservation of the recovery error \( \epsilon_{\text{GAI}} \) (either maximal or root mean square).

D. Calculation of reconstructed sample values

Based on the quantized post-interpolation residuals, the calculation of differential values is done, followed by the calculation of reconstructed sample values:
\[
\hat{R}_{\text{GAI}}(m, n) = Q_{\text{GAI}}(R, \epsilon_{\text{GAI}}),
\]
\[
\hat{I}_l(m, n) = \hat{I}_l(m, n) + \hat{R}_{\text{GAI}}(m, n).
\]

E. Statistical encoding

The quantized post-interpolation residuals undergo statistical encoding procedure.

III. SEMI-FRAGILE WATERMARKING FOR HGI COMPRESSION

The proposed watermarking system uses a hierarchical image representation and the HGI compression scheme with a changed quantizer.

Let \( I(m, n) \in [0, 255] \) be the source image, or cover image, \( B(k) \in [0, 1] \) be a binary sequence acting as a watermark. The correspondence between source image samples and the watermark is set by a certain mapping. As a result of this mapping, we obtain the following matrix:

\[
W(m, n) = F(R) = \begin{cases} 
-1, & \text{if } B(k) = 0 \\
1, & \text{if } B(k) = 1 
\end{cases}
\]

The mapping takes into account the hierarchical structure of the image used in HGI, and also uses a pseudo-random secret key, known both at the embedding and extraction side.

A. Interpolation

\[
\hat{I}_l(m, n) = P_{\text{GAI}} \left( \bigcup_{k=1}^{L-1} \{ \hat{I}_k(m, n) \} \right),
\]
where \( P_{\text{GAI}}(...) \) is the interpolation function.

B. Calculation of post-interpolation residuals

\[
R(m, n) = I_l(m, n) - \hat{I}_l(m, n).
\]

C. Quantization-based watermarking

\[
Q_{\text{GAI}}(R(m, n)) = Q_{\text{GAI}}(R, \epsilon_{\text{GAI}}),
\]
where \( Q_{\text{GAI}} \) is a quantizer based on QIM watermarking [22-23], and \( \epsilon_{\text{GAI}} \) is half of the QIM quantization step. This parameter determines the robustness of the watermark to additive white noise and the amount of distortion introduced by embedding.

D. Calculation of reconstructed sample values

\[
\hat{R}_{\text{GAI}}(m, n) = Q_{\text{GAI}}(R, \epsilon_{\text{GAI}}),
\]
\[
\hat{I}_l(m, n) = \hat{I}_l(m, n) + \hat{R}_{\text{GAI}}(m, n).
\]

When extracting the watermark from a received and possibly changed image \( \hat{I}^w(m, n) \), the same steps are performed, but watermark values \( \hat{W}(m, n) \) are restored at the quantization stage.

IV. EXPERIMENTAL INVESTIGATION OF THE PROPOSED WATERMARKING SYSTEM

When conducting the research, we held the following scheme. Suppose \( I \) be the source image, \( W \) be the watermark. Embedding \( W \) into \( I \) with parameter \( \epsilon_{\text{GAI}} \) is done according to Section 3. Distortions introduced by watermark embedding are estimated using the PSNR criterion, which indicates Peak Signal to Noise Ratio between the source image \( I \) and the watermarked image \( I^w \).

\( I^w \) can be subjected to any attacks, so on the receiver side, we have an image \( i^w \) that may differ from \( i^w \). The watermark \( \hat{W} \) is extracted from \( \hat{I}^w \) and compared with initial watermark \( W \). The paper considers two types of attacks: compression and local editing.

To carry out numerical experiments, ten images from the Waterloo Grayscale Set 1 and 2 [24] were chosen. Fragments of \( M = N = 257 \) size were extracted from each image. The matrix \( W \) was formed as follows: zero values for the samples of level \( l = L - 1 \) and the equally probable values 1 and -1 for other samples.

A. Specification of HGI and QIM parameters

The following parameter values were used.

- Bilinear interpolation functions as \( P_{\text{GAI}}(...) \) and \( P_{\text{GAI}}(...) \).

VI International Conference on “Information Technology and Nanotechnology” (ITNT-2020)
• HGI quantizer to ensure maximum recovery error $\varepsilon_{max}$:

$$Q_{\text{HGI}}(R, e_{\text{HGI}}) = e_{\text{HGI}} \cdot \text{Round}(\frac{R}{e_{\text{HGI}}}), e_{\text{HGI}} = 2^x \varepsilon, \ $$

$$Q^{-1}_{\text{HGI}}(R, e_{\text{HGI}}) = R_{\text{max}}. $$

• As a QIM quantizer, the simplest QIM family quantization function is used [23]:

$$Q_{\text{QIM}}(R, W, e_{\text{QIM}}) = \begin{cases} 2e_{\text{QIM}} \cdot \text{Round}(\frac{R}{2e_{\text{QIM}}}) + e_{\text{QIM}} \cdot \text{sign}(R - \text{Round}(\frac{R}{2e_{\text{QIM}}})), W = 1 \\ 2e_{\text{QIM}} \cdot \text{Round}(\frac{R}{2e_{\text{QIM}}}), W = 0, -1 \end{cases} $$

$$Q^{-1}_{\text{QIM}}(R_{\text{QIM}}, e_{\text{QIM}}) = R_{\text{QIM}}. $$

• Watermark extraction:

$$W(m, n) = 2 \cdot \text{mod}(\text{Round}(\frac{R_{\text{QIM}}}{e_{\text{QIM}}}), 2) - 1 $$

• HGI parameters setting the embedding rate: $l_{min}$, $l_{max}$ and $\theta$. Parameters $l_{min}$ and $l_{max}$ are the minimal and maximal hierarchical levels correspondingly, in which the embedding is performed. Parameter $\theta$ sets the percentage of current hierarchical level samples that undergo the watermark embedding.

B. Distortions introduced by watermarking

This subsection is aimed to investigate image distortions introduced by watermark embedding. Fig. 1 represents the dependence of $PSNR(I, I^w)$ on the value $e_{\text{QIM}}$ (the values were averaged for ten investigated images). The value $L = 7$ of the maximum number of hierarchical levels was used. The figure shows that the $PSNR$ value decreases with increasing $e_{\text{QIM}}$.

![Fig. 1. Dependence of $PSNR(I, I^w)$ on $e_{\text{QIM}}$ value.](image)

The next part of the investigation is concerned with the question of how the embedding applied to a subset of hierarchical levels affects on distortions of the source image. Table 1 gives the $PSNR$ values for different combinations of $l_{min}$ and $l_{max}$ parameters values (averaged for investigated images). The table shows that the more samples undergo the watermarking, the lower $PSNR$ value is. $PSNR$ value has achieved the minimum for embedding in all hierarchical levels ($l_{min} = 0$, $l_{max} = 7$) and the maximum for embedding only in the highest hierarchical level ($l_{min} = l_{max} = 7$).

<table>
<thead>
<tr>
<th>$l_{max}$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_{min}$</td>
<td>0.2765</td>
<td>0.3095</td>
<td>0.3059</td>
<td>0.3050</td>
<td>0.3047</td>
<td>0.3048</td>
<td>0.3047</td>
<td>0.2685</td>
</tr>
<tr>
<td>1</td>
<td>0.3265</td>
<td>0.3247</td>
<td>0.3244</td>
<td>0.3242</td>
<td>0.3244</td>
<td>0.3243</td>
<td>0.3243</td>
<td>0.2591</td>
</tr>
<tr>
<td>2</td>
<td>0.3333</td>
<td>0.3326</td>
<td>0.3325</td>
<td>0.3322</td>
<td>0.3323</td>
<td>0.3334</td>
<td>0.3334</td>
<td>0.2683</td>
</tr>
<tr>
<td>3</td>
<td>0.3354</td>
<td>0.3352</td>
<td>0.3352</td>
<td>0.3353</td>
<td>0.3353</td>
<td>0.3354</td>
<td>0.3354</td>
<td>0.2683</td>
</tr>
<tr>
<td>4</td>
<td>0.3361</td>
<td>0.3360</td>
<td>0.3359</td>
<td>0.3359</td>
<td>0.3359</td>
<td>0.3359</td>
<td>0.3359</td>
<td>0.2683</td>
</tr>
<tr>
<td>5</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
</tr>
<tr>
<td>6</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
</tr>
<tr>
<td>7</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
<td>0.3363</td>
</tr>
</tbody>
</table>

One more parameter that affects the degree of the distortion is the percentage of embedding into each hierarchical level $\theta$. Fig. 2 represents the dependence of $PSNR$ on $\theta$ value for the fixed $e_{\text{QIM}} = 20$. Again, the more samples are used for embedding, the lower $PSNR$ value is.

So, we can make the conclusion that the degree of cover image distortions, resulting from watermark embedding, can be regulated by changing $e_{\text{QIM}}$, $l_{min}$, $l_{max}$, and $\theta$.

![Fig. 2. Dependence of $PSNR$ on the embedding percentage $\theta$ (for $e_{\text{QIM}} = 20$).](image)

C. Watermark restoration after a compression attack

In this experiment, the watermarked image $I^w$ is subjected to HGI compression with $e_{\text{HGI}}$, and $I^w$ is the compressed image. Denote $W$ the extracted watermark (using the parameter $e_{\text{HGI}}$). In this subsection, we are interested in how the ratio of HGI and QIM parameters affects the restoration of the watermark, as well as what distortions the embedding and compression introduce into the cover image.

For $W$ and $W$ comparison, the BER (Bit Error Rate) criterion was used. Fig. 5 shows the dependence of $BER(W, W)$ and maximal deviation $e_{max}$ on $e_{HGI}$ under the fixed value $e_{QIM}$ (averaged values for ten investigated images).
Investigations show that BER is zero for values 
\[ \varepsilon_{\mu_{GI}} \leq \frac{\varepsilon_{QIM}}{2}, \]
as well as for \[ \varepsilon_{\mu_{GI}} = \varepsilon_{QIM}. \]
When \[ \varepsilon_{\mu_{GI}} > \varepsilon_{QIM} \] BER undergoes a sharp jump and is set at a value of about 0.5 (which indicates the destruction of the watermark).

When \[ \varepsilon_{\mu_{GI}} = \varepsilon_{QIM} \], the watermark is extracted without errors, since after the embedding, all interpolation residuals are multiple \[ \varepsilon_{QIM} \], and their re-quantization in HGI on the same value makes no changes.

The maximum deviation between initial cover image and the image reconstructed after embedding and compression equals

\[ \varepsilon_{\text{max}} = \max \| \hat{I} - \tilde{I} \| = \begin{cases} \varepsilon_{QIM}, \varepsilon_{\mu_{GI}} \text{ is divider of } \varepsilon_{QIM}, \\ \varepsilon_{QIM} + [0.5 \cdot \varepsilon_{\mu_{GI}}], \text{ otherwise.} \end{cases} \]

**D. Reconstruction of local distortions area**

In this subsection we analyze the accuracy of local distortion area estimation. To model local distortions, for simplicity, we replaced samples within a predefined mask by random samples, as shown in Fig. 6. Such processing makes it easy to estimate the mask of distortions without using a watermark, but we have not used any information on the type of distortions in the investigation. We will denote the resulting image as \( \tilde{I} \).

On the receiving side \( \hat{W} \) is extracted from \( \hat{I} \) and is compared with \( W \) that is generated using the same secret key as on the embedding side (with the same values \( l_{\text{max}} \), \( l_{\text{max}} \), and \( 0 \)). The difference between \( \tilde{W} \) and \( W \) is not equal to \( D \) (because watermark bits are correctly extracted from approximately half of the distorted samples randomly) (see Fig. 6).

Semi-fragile watermarking is always a compromise between the watermark robustness and imperceptibility. In the case of watermark embedding in selective hierarchical levels and with \( \theta < 100 \% \) we have got a sparse structure of nonzero difference between \( W \) and \( \hat{W} \). The following algorithm is suggested for local distortions area reconstruction \( \tilde{D} \).

In the **first step**, going down from the samples of the highest hierarchical level to \( l_{\text{max}} \), when \( W(m,n) \neq W(m,n) \) and \( W(m,n) = 0 \) we fill with "1" its neighborhood \( \tilde{D}(m \pm (2^l - 1), n \pm (2^l - 1)) = 1 \), where \( l \) is a current level. This area can "suffer" from the distortion of sample \( (m,n) \) at the sequential reconstruction of samples.
In the second step, we process $D$ with two sequential local morphological filters (max-min) with window size $w_{in} = 2 \cdot 2^{m-1} + 1$.

To investigate the quality of the proposed algorithm, we use the following criteria: $q_{01}$, the relative quantity of false positive detections of local distortion area; $q_{10}$, the relative quantity of omissions, their sum $q$:

$$q_{01} = \frac{\big| D = 0, \hat{D} = 1 \big|}{\big| D = 1 \big|}, \quad q_{10} = \frac{\big| D = 1, \hat{D} = 0 \big|}{\big| D = 1 \big|},$$

where $\big| \big|$ is cardinality of a set. We define denominator as the size of local distortions area to make criteria values comparable for big and small areas.

The next part of the current subsection is aimed to investigate how criteria values change for different values of $l_{\text{min}}, l_{\text{max}}, \text{and } \theta$. In this part, we use only the "Lena" image, and all presented numerical values have been averaged over 100 observations.

Fig. 7 presents the dependence of $\text{PSNR}(I, I''')$ on $\theta$ for different combinations of $l_{\text{min}}, l_{\text{max}}$, and fixed value $\epsilon_{\text{QIM}} = 20$. The figure shows that for $l_{\text{min}} = l_{\text{max}} = 0$, the curve lies too low (cover image distortions are too evident). Watermark embedding in the second hierarchical level provides better PSNR, but $q$ is too high. So, the cases $l_{\text{min}} = l_{\text{max}} = 1$ and $l_{\text{min}} = 1, l_{\text{max}} = 2$ should be analyzed.

Fig. 8 shows curves for $q, q_{01}, q_{10}$ depending on $\theta$ for $l_{\text{min}} = l_{\text{max}} = 1$ and $l_{\text{min}} = 1, l_{\text{max}} = 2$ for a fixed value $\epsilon_{\text{QIM}} = 20$. The comparison can be done as follows. Suppose we are interested in the value of the relative quantity of omissions $q_{10} = 0.05$, we can find the intersection with $q_{10}$ curve and get $\theta$ and $q$ values (see green dashed line and values). So, we get that for $l_{\text{min}} = l_{\text{max}} = 1$ the criterion value $q=0.2$, and for $l_{\text{min}} = 1, l_{\text{max}} = 2$ the criterion value $q=0.3$. Therefore, values $l_{\text{min}} = l_{\text{max}} = 1$ and $\theta=70..100\%$ can be recommended for the embedding in the case of local distortion attack.

Fig. 9-10 represent some results of the proposed algorithm of local distortions area reconstruction for parameter values, highlighted in Fig. 8 (see green circle). The comparison shows better results for Fig. 9 than for Fig. 10.

V. CONCLUSIONS

In this paper, we have proposed a semi-fragile watermarking system adopted for the HGI compression algorithm. Its main idea is to utilize the HGI scheme in the watermarking procedure and to replace the quantization of interpolation residuals with watermark embedding using a QIM-based method. This approach makes it possible to obtain the robustness against HGI compression in a predefined range of quality factors.

The parameters of the proposed algorithm allow us to find the compromise between distortions, contributed by the embedding, and robustness to certain attacks.

The conducted experiments have shown the ability of the proposed watermarking system to protect images with high quality in terms of PSNR. We also investigated the accuracy of local distortion detection. As a result, a trade-off between image quality and forgery detection accuracy has been found.

Future work may include the investigation of some other QIM family quantizers, including those providing fewer distortions (like DC-QIM, distortion compensated QIM [22]) and more protected ones (like IM-QIM, statistically immune QIM [23]).

ACKNOWLEDGMENT

The work was funded by the RSF grant #18-71-00052 (in parts of watermarking system construction and investigation), and by the RFBR grant #19-29-09045 (in part of application and parameters adaptation to protect remote sensing data and local distortion estimation).
REFERENCES


