Content-adaptive reliable robust lossless data embedding

Lingling An\textsuperscript{a}, Xinbo Gao\textsuperscript{a,b}, Yuan Yuan\textsuperscript{c,*}, Dacheng Tao\textsuperscript{d}, Cheng Deng\textsuperscript{a}, Feng Ji\textsuperscript{a}

\textsuperscript{a} VIPS Lab, School of Electronic Engineering, Xidian University, Xi’an 710071, China
\textsuperscript{b} Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi’an 710071, China
\textsuperscript{c} State Key Laboratory of Transient Optics and Photonics, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an 710119, China
\textsuperscript{d} Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Broadway NSW 2007, Australia

A R T I C L E  I N F O

Article history:
Received 19 February 2011
Received in revised form 24 June 2011
Accepted 26 August 2011
Communicated by M. Wang
Available online 18 September 2011

Keywords:
Histogram rotation
Just noticeable distortion
Robust lossless data embedding

A B S T R A C T

It is well known that robust lossless data embedding (RLDE) methods can be used to protect copyright of digital images when the intactness of host images is highly demanded and the unintentional attacks may be encountered in data communication. However, the existing RLDE methods cannot be applied satisfactorily to the practical scenarios due to different drawbacks, e.g., serious “salt-and-pepper” noise, low capacity and unreliable reversibility. In this paper, we propose an effective solution to RLDE by improving the histogram rotation (HR)-based embedding model. The proposed method is a content-adaptive reliable RLDE or CAR for short. It eliminates the “salt-and-pepper” noise in HR by the pixel adjustment mechanism. Therefore, reliable regions for embedding can be well constructed. Furthermore, we basically expect the watermark strengths to be adaptive to different image contents, and thus we have a chance to make an effective tradeoff between invisibility and robustness. The luminance masking together with the threshold strategy is duly adopted in the proposed RLDE method, so the just noticeable distortion thresholds of different local regions can be well utilized to control the watermark strengths. Experimental evidence on 300 test images including natural, medical and synthetic aperture radar (SAR) images demonstrates the effectiveness of the proposed data embedding method.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Recently, lossless data embedding (LDE) has attracted more and more attention for information security problems in multimedia processing [2,6,12,14]. It deals with the situation where the intactness of host images after the extraction of the watermarks (e.g., copyright or authentication messages) is highly demanded. A dozen practical problems fall into this situation because the unintentional attacks, e.g., random noise and lossy compression, often degrade the quality of the watermarked images during the transmission, leading to a failure of the extraction of the hidden watermarks [1]. As a consequence, it is necessary to reconsider LDE methods by taking the robustness against the attacks into account.

Many robust lossless data embedding (RLDE) methods have been developed in recent years. For example, De Vleeschouwer et al. [8,9] adopted the center of mass as a discriminating factor, and at the same time rotated the vectors pointing from the center of a circle to the center of mass in opposite ways to embed the watermarks. Correspondingly, the sign of the smallest angle between the above vectors is employed to extract the hidden watermarks at the receiver side. Because both the center of mass and the rotation of the vectors are closely related to the histogram distribution, we term this kind of method as the histogram rotation (HR)-based RLDE [3]. Although HR-based RLDE has achieved good robustness against JPEG compression, the “salt-and-pepper” noise greatly degrades the quality of the watermarked images. Moreover, extensive experimental results show that this noise sometimes leads to obstacles to recovering the host images and extracting the watermarks. Aiming to eliminate the “salt-and-pepper” noise in this kind of data embedding method, the histogram distribution constrained (HDC) methods have been developed in spatial and wavelet domains [19,20,25,26], respectively, which are based on selectively modifying the particular statistical quantities to embed the watermarks with the help of the block classification strategy and error correction coding (ECC). Unfortunately, the HDC suffers from low capacity and unreliable reversibility [3]. The new block selection mechanism and parameter model are exploited thereafter to improve its reversibility in [13]. However, the problem of low capacity remains unsolved. In summary, the HDC methods are free of “salt-and-pepper” noise and achieve high invisibility at the cost of capacity. As for the robustness, there is no remarkable improvement in comparison with HR-based methods. In view of this, a natural concern is the possibility of searching for another effective solution for “salt-and-pepper” noise in HR to achieve better comprehensive performance from the aspects of capacity, invisibility, and robustness.
In this paper, we tackle this problem by constructing a content-adaptive reliable embedding scheme or CAR for short. In particular, the pixel adjustment (PA) mechanism [24] is employed to solve the “salt-and-pepper” noise problem. However, it is not optimal for embedding watermarks into different images. This is because we consider neither the just noticeable distortion (JND) thresholds of different regions in an image nor the tradeoff between invisibility and robustness, though both of them are important for RLDE. As a consequence, to further improve the performance, luminance masking [7] together with the threshold strategy, is used to estimate the JND thresholds of different regions, which determine the watermark strengths in the host images, and thus balance visibility and robustness. To justify the effectiveness of the proposed CAR data embedding method, we compare it with the typical HR-based and HDC methods [20, 26] and the experimental results show its superiority in comprehensive performance including invisibility, capacity, and robustness.

The rest of the paper is organized as follows. In Section 2, the visual perceptual models, especially luminance masking, are briefly presented for readers to better understand the proposed method. Section 3 details the proposed method from the following four aspects: blocking, embedding level computation, pixel adjustment and embedding based on rotation. In Section 4, thorough experimental results conducted on natural, medical and synthetic aperture radar (SAR) images are presented for performance evaluation. The final section concludes the paper.

2. Visual perceptual model

The perceptual characteristics of human visual system (HVS) play an important role in many practical applications, e.g., image quality assessment [17], scene classification [21], and digital watermarking [18]. For image watermarking, a key issue is to find the perceptually unimportant image regions to embed the watermarks. To this end, many visual perceptual models have been applied to the watermarking methods, e.g., Barni et al.’s pixel-wise masking [5], Watson’s JND model [22], and Chou et al.’s luminance masking [7]. Because the first two models were developed in transform domains, we use the luminance masking to estimate the JND thresholds of local regions in this paper.

Luminance masking proposed by Chou [7] incorporates the properties of HVS into the estimation of JND thresholds to measure the perceptual redundancies in an image. It considers two factors: the average background luminance and its non-uniformity, and can be used to control the watermark strengths in specific watermarking embedding methods. Consider a pixel at \((i, j)\) in a given image with the size of \(H \times W\), the JND threshold is described by

\[
\text{JND}(i,j) = \max [f_s(l_{bg}(i,j))f_s(l_{mg}(i,j)), l_{mg}(i,j)]
\]

where

\[
f_s(l_{bg}(i,j)) = \begin{cases} (3/128) \cdot l_{bg}(i,j) - 127 & l_{bg}(i,j) > 127 \\ 17 \cdot (1 - \sqrt{l_{bg}(i,j)/127}) + 3, l_{bg}(i,j) \leq 127 \end{cases}
\]

and

\[
f_v(l_{bg}(i,j),l_{mg}(i,j)) = l_{mg}(i,j)d_1(l_{bg}(i,j)) + d_2(l_{bg}(i,j)),
\]

represent the visibility threshold and spatial masking functions, respectively. In Eq. (3), the background luminance dependent functions \(d_1(l_{bg}(i,j))\) and \(d_2(l_{bg}(i,j))\) are defined by

\[
d_1(l_{bg}(i,j)) = 0.0001 \times l_{bg}(i,j) + 0.115,
\]

and

\[
d_2(l_{bg}(i,j)) = 0.5 - 0.01 \times l_{bg}(i,j).
\]

Here \(l_{mg}(i,j)\) calculates the maximum weighted average of luminance changes around the pixel at \((i,j)\) in four directions, defined by

\[
l_{mg}(i,j) = \max_k (|g_k(i,j)|),
\]

\[
g_k(i,j) = \frac{1}{16} \sum_{\Delta l = 1}^{5} \sum_{\Delta j = 1}^{5} p(i - 3 + \Delta l, j - 3 + \Delta j) \cdot G_k(i, j),
\]

where \(p(i,j)\) is the pixel at \((i,j)\) and \(G_k(i, j)\) denotes the operator in the \(k\)th direction. In addition, the average background luminance, denoted by \(b_{bg}(i,j)\), is computed by

\[
b_{bg}(i,j) = \frac{1}{32} \sum_{\Delta l = 1}^{5} \sum_{\Delta j = 1}^{5} p(i - 3 + \Delta l, j - 3 + \Delta j) \cdot K(i, j),
\]

and \(K(i, j)\) is a weighted low-pass operator. In this model, the definitions of five operators are shown in Fig. 1.

3. The proposed method

In this section, we introduce a novel RLDE method, i.e., CAR, which finds an effective solution to “salt-and-pepper” noise and improves the comprehensive performance of HR-based data embedding methods. Fig. 2 shows the proposed general framework of CAR for robust lossless data embedding. It first builds the non-overlapping blocks in an image. Based on the blocks, the luminance masking along with the threshold estimation is performed to get the optimal watermark strengths. Afterwards, the grayscale values of the pixels likely to overflow or underflow during watermark embedding are changed by the pixel adjustment mechanism. Finally, the embedding model based on histogram rotation is employed to derive the watermarked image. In the next sub-sections, we will present four components of the proposed framework in detail.

For convenience, Table 1 lists important symbols used in watermark embedding. In addition, superscript characters such as \(w, r\) and ‘ are utilized to represent the watermarked, recovered and reliable items corresponding to the host ones, respectively. For example, \(p^w\) and \(\Gamma\) are the watermarked and recovered ones of the host image \(I\).

3.1. Blocking

Empirically, the blocking scheme is helpful to capacity control and robustness against the unintentional attacks, e.g., JPEG compression.
incorporating the perceptual properties of the HVS and threshold to balance robustness and invisibility is another important issue. Moreover, how embedding level is applied to all of the blocks in a host image is irrelevant to the image content. That is to say, the same capability and robustness. In most RLDE methods, the embedding level directly determines the change of pixels, and thus affects invisibility. In this case, the visual quality of the watermarked images is too poor for practical scenarios although high robustness is achieved. To handle this problem, we introduce a threshold strategy to further optimize the embedding level of each block, and thus balance robustness and invisibility. Let $L_{B_k}$ be the embedding level of the block $B_k$, then it can be calculated by

$$L_{B_k} = \begin{cases} T, & J_{B_k} > T \\ I_{B_k}, & \text{otherwise} \end{cases},$$

where $T$ is a threshold. By Eq. (11), an optimal embedding level can be obtained by removing those JND thresholds greater than $T$ in a host image, which will be applied to the pixel adjustment and watermark embedding in the next sub-sections.

### 3.3. Pixel adjustment

As aforementioned, the serious “salt-and-pepper” noise has been deemed as an obvious drawback of HR-based RLDE methods. Although the HDC methods have partially solved this problem, the loss of capacity and unstable reversibility may lead to obstacles to their practical applications. Therefore, motivated by the successes of the PA mechanism for handling overflow and underflow of pixels [24], we combine it with the above content-adaptive scheme and build a content-adaptive reliable embedding framework.

Given the $k$-th block $B_k$ with the size of $H \times W$ in a $t$-bit image and its embedding level $L_{B_k}$, we define two unreliable sets as

$$S_u = \{B_k^{(i)} | B_k^{(i)} > 2^t - 1 - Q_k\},$$

and

$$S_l = \{B_k^{(i)} | B_k^{(i)} < Q_k\},$$

in which $B_k^{(i)}$ is the pixel at $(i,j)$, $1 \leq i \leq h$, $1 \leq j \leq w$, and $Q_k$ is the reliable level. The key of PA is to modulate the grayscale values of the pixels belonging to the unreliable sets into a reliable range, i.e., $[Q_k, 2^t - 1 - Q_k]$, so as to prevent “salt-and-pepper” noise. Here, we define $Q_k > L_{B_k}$ in order to make pixel adjustment adaptive to the embedding level, which is to say, the PA mechanism is always effective for any embedding level. At the same time, $Q_k = L_{B_k} + 1$ is the most appropriate choice by taking invisibility into account. Let $S_u$ and $S_l$ be the reliable results of $S_u$ and $S_l$ after PA, then the adjusted pixels can be obtained by Eqs. (12) and (13), denoted as

$$S_u^{(i)} = S_u^{(i)} - Q_k,$$

$$S_l^{(i)} = S_l^{(i)} + Q_k.$$

### Table 1

Important symbols used in watermark embedding.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
<td>Given host image</td>
</tr>
<tr>
<td>$H \times W$</td>
<td>Image size</td>
</tr>
<tr>
<td>$B_k$</td>
<td>k-th block</td>
</tr>
<tr>
<td>$J_{B_k}^{(i)}$</td>
<td>JND threshold of $B_k^{(i)}$</td>
</tr>
<tr>
<td>$T$</td>
<td>Threshold</td>
</tr>
<tr>
<td>$L_{B_k}$</td>
<td>Embedding level of $B_k$</td>
</tr>
<tr>
<td>$m_k$</td>
<td>k-th bit of $M$</td>
</tr>
<tr>
<td>$M_k$</td>
<td>Random sets of $B_k$</td>
</tr>
<tr>
<td>$S_{u/l}$</td>
<td>Unreliable sets of $B_k$</td>
</tr>
<tr>
<td>$Q_{k}$</td>
<td>Watermark sequence</td>
</tr>
<tr>
<td>$S_{u/l}$</td>
<td>Centroid vectors of $Z_{u/l}$</td>
</tr>
</tbody>
</table>

Both of the aforementioned two kinds of RLDE methods, i.e., HR and HDC, utilized this scheme. Considering a host image with the size of $H \times W$, we also divide it into non-overlapping blocks, denoted by

$$\Omega = \left\{ B_k | k = 1, 2, \ldots, \left[ \frac{H}{h} \right] \times \left[ \frac{W}{w} \right] \right\},$$

where $h \times w$ is the block size. In the embedding, each block is associated with a bit of the watermark sequence $M = [m_1, m_2, \ldots, m_n]$, which is to say, the watermarks will be embedded into the blocks one by one in a specific order.

### 3.2. Embedding level computation

The embedding level denotes the watermark strength, which directly determines the change of pixels, and thus affects invisibility and robustness. In most RLDE methods, the embedding level is irrelevant to the image content. That is to say, the same embedding level is applied to all of the blocks in a host image. However, it is not optimal because the differences of JND thresholds among the blocks are not taken into account. Moreover, how to balance robustness and invisibility is another important issue to be considered in robust watermarking methods. To target these two problems, we propose the content-adaptive embedding by incorporating the perceptual properties of the HVS and threshold strategy. To be specific, given the $k$-th block $B_k$, the luminance masking, a typical JND model in spatial domain, is utilized to estimate the JND threshold of $B_k$, denoted by

$$J_{B_k} = \frac{1}{H \times W \times \sum_{i=1}^{h} \sum_{j=1}^{w} J_{B_k}^{(i,j)}}.$$  

Here $J_{B_k}^{(i,j)}$ calculates the JND threshold of a pixel at $(i,j)$ in the block $B_k$ by Eq. (1). Based on the local JND thresholds of different blocks, Eq. (10) finds a more optimal embedding scheme by setting a suitable embedding level for each block. In practical applications, however, this scheme does not always take a good tradeoff between robustness and invisibility into consideration, leading to somewhat low visual quality for some watermarked images. For example, because the adjacent pixels in medical images are more closely correlated, the JND thresholds of some blocks can be up to 20, which determine high embedding levels for these blocks. In this case, the visual quality of the watermarked images is too poor for practical scenarios although high robustness is achieved. To handle this problem, we introduce a threshold strategy to further optimize the embedding level of each block, and thus balance robustness and invisibility. Let $L_{B_k}$ be the embedding level of the block $B_k$, then it can be calculated by

$$L_{B_k} = \begin{cases} T, & J_{B_k} > T \\ I_{B_k}, & \text{otherwise} \end{cases},$$

where $T$ is a threshold. By Eq. (11), an optimal embedding level can be obtained by removing those JND thresholds greater than $T$ in a host image, which will be applied to the pixel adjustment and watermark embedding in the next sub-sections.
where \( S_{A}^{k} \in S_{A} \) and \( S_{B}^{k} \in S_{B} \). Thus, a reliable block can be obtained, and thereafter it will be used for watermark embedding. To further show the effectiveness of the PA mechanism, an example is given below to illustrate the above procedure. Given \( L_{h} = 8 \) and \( Q_{w} = 9 \), we consider an arbitrary pixel with the grayscale value of 250 in an 8-bit host image. Since the pixel belongs to \( S_{A} \), its grayscale value needs to be changed from 250 to 241 by Eq. (14) so that it will not be greater than 255 even if the \( L_{h} \) is added to it when embedding “1”. That is to say, the overflow of pixels can be avoided. In this way, the “salt-and-pepper” noise existing in HR-based methods can be eliminated successfully. Moreover, the problems of capacity and reversibility in HDC are also solved effectively using the PA mechanism, which will be demonstrated in Section 4.

### 3.4. Embedding based on rotation

In this sub-section, we consider the watermark embedding task and show how to implement the CAR framework based on histogram rotation. Taking the \( k \)-th block \( B_{k} \) as an example, we summarize the embedding process as the following four stages, i.e., zone separation, histogram mapping, centroid computation and vector rotation. To be specific, two random zones with the same size, denoted by \( Z_{A} \) and \( Z_{B} \), are firstly separated from \( B_{k} \). To make the histogram distributions of two zones as similar as possible, we define \( Z_{A} \) and \( Z_{B} \) by

\[
Z_{A} = \left\{ B_{k}^{(i,j)} | B_{k}^{(i,j)} \in B_{k}, \mod(i+j,2) = 0, 0 \leq i \leq h, 0 \leq j \leq w \right\},
\]

and \( Z_{B} = Z_{A} \), i.e., \( Z_{B} \) is the complement of \( Z_{A} \). Secondly, the histograms of \( Z_{A} \) and \( Z_{B} \) are mapped into the circles, as shown in Fig. 3, wherein an 8-bit image is considered. In particular, the positions scattering over the circle, denoted by small circles, correspond to the grayscale values of the pixels, and a weight proportional to the frequency of each grayscale value is assigned to each position. In Fig. 3, given a position \( i \), the magnitude of the vector pointing to \( i \), i.e., \( |\overrightarrow{V}_i| \), is used to denote the weight of \( i \). Next, the centers of mass of the zones are computed. Let \( M_{A} \) be the center of mass of \( Z_{A} \), then its coordinates \((x_{A}, y_{A})\) can be computed by

\[
x_{A} = \frac{\sum_{i=1}^{255} |\overrightarrow{V}_i| \cos\omega_i}{\sum_{i=1}^{255} |\overrightarrow{V}_i|}, \quad y_{A} = \frac{\sum_{i=1}^{255} |\overrightarrow{V}_i| \sin\omega_i}{\sum_{i=1}^{255} |\overrightarrow{V}_i|},
\]

in which \( \omega_i \) represents the angle of \( \overrightarrow{V}_i \). Similar to Eq. (16), the center of mass of \( Z_{B} \), \( M_{B} \), can also be computed. Finally, we take \( \overrightarrow{V}_{M_A} \) and \( \overrightarrow{V}_{M_B} \) as the vectors pointing from the center of the circle to \( M_{A} \) and \( M_{B} \), and embed a bit of watermarks into the block \( B_{k} \) by slightly rotating \( \overrightarrow{V}_{M_A} \) and \( \overrightarrow{V}_{M_B} \) in opposite directions. For instance, \( \overrightarrow{V}_{M_B} \) is rotated clockwise and \( \overrightarrow{V}_{M_A} \) anticlockwise to embed bit “1”, as shown in Fig. 2. To simplify the notation, we call \( \overrightarrow{V}_{M_A} \) and \( \overrightarrow{V}_{M_B} \) the centroid vectors in the rest of this paper. By applying the aforementioned stages to all of the blocks, the watermark sequence can be embedded into the host image. It should be noted that the embedding procedure assumes the block is non-problematic so that a reversible recovery at the receiver side can work successfully. Correspondingly, if a block is determined to be problematic, no watermark bit is embedded into it; at the same time, additional knowledge about its index is stored. Refer to [9] for the definition of problematic blocks.

According to the aforementioned descriptions, we can summarize the embedding procedure of the proposed CAR, depicted in Table 2. At the receiver side, the extraction procedure is similar to the embedding one. In particular, the watermarked image is firstly divided into non-overlapping blocks. Following that, the zone separation and centroid computation are applied to each block. Based on this, one can extract the hidden watermark sequence by checking the sign of the small angle between centroid vectors. Finally, the inverse histogram rotation and PA mechanism are utilized to recover the image. Table 3 illustrates the extraction procedure of CAR.

### 4. Experimental results and analysis

In this section, we justify the effectiveness of the proposed content-adaptive reliable robust lossless data embedding, i.e., CAR, on natural, medical and SAR images in comparison with three typical RLDE methods, e.g., HR [9], HDC (1) [20] and HDC (2) [26]. In addition, in order to make a comprehensive comparison, the visibility, capacity and robustness are utilized to evaluate the performance of the methods. In particular, the lossy compression,
Table 3

<table>
<thead>
<tr>
<th>Extraction procedure of CAR method.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> A watermarked image ( I' ) with the size of ( H \times W ), the block size ( h \times w ), and the embedding levels ( L = [l_1, l_2, \ldots, l_n] ).</td>
</tr>
<tr>
<td><strong>Output:</strong> The recovered watermark sequence ( M' = [m'_1, \ldots, m'_n] ) and image ( I' ).</td>
</tr>
</tbody>
</table>

1. Divide \( I' \) into \( n \) non-overlapping blocks, \( \{B_k^i|k = 1, \ldots, H|W|/h\} \).
2. For \( k = 1 \) to \( n \) do
   
   1. Perform zone separation (ZS) to get two zones of block \( B_k^i \)
      \[ Z_k^H, Z_k^L \leftarrow ZS(B_k^i) \]
   2. Compute the centroid vectors of \( Z_k^H \) and \( Z_k^L \), denoted as \( \bar{v}_{Z_k^H} \) and \( \bar{v}_{Z_k^L} \), respectively.
   3. Perform the watermark extraction procedure \( W \)
      \[ L_k = \begin{cases} 0, & \text{if } \alpha_k > 0 \\ 1, & \text{otherwise} \end{cases} \]
   4. Perform the image recovery by HR \( W \)
      \[ B_k^i \leftarrow HR(B_k^i, m'_k, \bar{v}_{Z_k^H}) \]
   5. Perform PA for unreliable blocks to get the recovered block \( B_k^i \)
      \[ B_k^i \leftarrow PA(B_k^i, Q_k) \]
3. End for
4. Return the recovered watermark sequence \( M' = [m'_1, \ldots, m'_n] \) and image \( I' \)

i.e., JPEG and JPEG2000 compression, and additive Gaussian noise (AGN) are applied to the watermarked images to test the robustness. Here, the quality factor of JPEG compression ranges from 20 to 100 with a step of 10, the bit rate of JPEG2000 is from 0.2 to 2.0 with a step of 0.2 where the Kakadu command line tool is adopted, the mean of AGN is zero and the variance changes from 0.01 to 0.04 with a step of 0.01. Besides, the robustness, the peak signal-to-noise ratio (PSNR) is employed to estimate the distortion of the watermarked images versus the host ones, i.e., invisibility. Given a host image \( I \) and its watermarked one \( I'' \) with the size of \( H \times W \), PSNR is defined as

\[
\text{PSNR} = 10 \log \left( \frac{255^2 \times H \times W}{\sum_{i,j}^H \sum_{j=1}^W (I(i,j) - I''(i,j))^2} \right),
\]

where \( l(i,j) \) and \( l''(i,j) \) denote the grayscale values of the pixels at \((i,j)\) in the host and watermarked images. Meanwhile, the pure capacity expressed in bits/pixel is computed to show how many watermarks can be embedded into a host image.

4.1. Dataset

To demonstrate the effectiveness and applicability of CAR, we build the dataset with three kinds of images, i.e., 100 natural images selected from the CVG-UGR image database [27], 100 medical images and 100 SAR images. In particular, medical images are constructed based on DICOM sample image sets [28] and Osirix website [29], including CT and MR images. SAR images are picked up from some open image database websites. In our experiments, all images are set to a fixed size, i.e., \( 512 \times 512 \times 8 \) for natural and medical images and \( 256 \times 256 \times 8 \) for SAR images so as to facilitate the comparison.

4.2. Parameter analysis

As discussed in Section 3, the parameters in the proposed CAR greatly affect its performance. In this sub-section, we will pay more attention to the block size, threshold and their significant effects on performance, i.e., the effect of block size on capacity, and effects of threshold on invisibility and robustness.

4.2.1. Block size

As known, the blocking scheme is helpful to capacity control. The smaller the block is, the higher the capacity is, and vice versa. Specifically, the pure capacity of CAR is dependent on the number of problematic blocks and the length of the additional knowledge, similar to HR. Given an image \( I \) with the size of \( H \times W \) and the block size \( h \times w \), the pure capacity expressed in bits/pixel, denoted by \( C_p \), can be computed by

\[
C_p = \frac{|\Omega| - |P| - |S|}{H \times W},
\]

where \( |\Omega| = |H/h| \times |W/w| \) denotes the total number of blocks, \( |P| \) is the number of problematic blocks, and \( |S| \) means the length of the additional knowledge. Because \( |S| \) is somewhat small in practical scenarios in comparison with others, we only consider the effects of \( |\Omega| \) and \( |P| \) on \( C_p \) here.

Table 4 shows the average results of problematic blocks with different block sizes for the three kinds of test images, wherein, the threshold is 8. As shown in Table 4, with the increase of the block size, \( |P| \) is becoming smaller and smaller, close to 0 bit for the block size 16 × 16. This is because the adjacent pixels in a block are more closely correlated for big blocks than for small blocks. That is to say, the centroid vectors of two random zones \( Z_h \) and \( Z_l \) in a block, i.e., \( \bar{v}_{Z_h} \) and \( \bar{v}_{Z_l} \), are closer to each other, which will be helpful to deducing the number of the problematic blocks according to [9].

Based on the results shown in Table 4, the pure capacity \( C_p \) can be approximately represented by \( C_p \approx |\Omega|/(H \times W) \), i.e., it is inversely proportional to the block size \( H \times W \). Fig. 4 presents the statistical average of \( C_p \) with the threshold of 8 and the block size from 4 × 4 to 16 × 16 with a step of 4, which further demonstrates the significant effects of block size on pure capacity. By contrast, the block size has no significant effect on invisibility because almost all of the pixels are changed when embedding for different block sizes.

4.2.2. Threshold

Apart from the block size, the threshold \( T \), as another important parameter in CAR, optimizes the embedding level and thus affects invisibility and robustness. In this part, we will make an analysis of the effects of \( T \) on these two aspects. In our experiments, the block size is fixed at 8 × 8, \( T \) is from 6 to 12 with a step of 1, and we suppose all of the blocks in the host image are non-problematic.

Let \( D(i,j) \) be the absolute value of the difference in grayscale values between the pixels at \((i,j)\) in the host and watermarked images, i.e., \( D(i,j) = |l(i,j) - l''(i,j)| \). There are three values of \( D(i,j) \) according to the embedding rules of CAR discussed in Section 3. That is

\[
D(i,j) = \begin{cases} D_1 = L_h, & \text{if } l(i,j) \geq l''(i,j) \\ D_2 = Q_h + L_h, & \text{if } l(i,j) < l''(i,j) \text{ and } l(i,j) - l''(i,j) \geq 0 \\ D_3 = Q_h - L_h, & \text{otherwise} \end{cases}
\]

Table 4

<table>
<thead>
<tr>
<th>Number of problematic blocks with different block sizes.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Image</strong></td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Natural</td>
</tr>
<tr>
<td>Medical</td>
</tr>
<tr>
<td>SAR</td>
</tr>
</tbody>
</table>

L. An et al. / Neurocomputing 79 (2012) 1–11
in which \( D(i,j) = D_1 \) if the pixel at \((i,j)\) is unadjusted by the PA mechanism before embedding; and \( D(i,j) = D_2 \) if PA is used and the direction of adjustment is the same as that of the embedding procedure, for example, both of them increase the value of \( D(i,j) \). If the PA and embedding procedure modify the value of \( D(i,j) \) in an opposite way, \( D(i,j) = D_3 \). According to Eq. (19), Eq. (17) can be rewritten as

\[
\text{PSNR} = 10 \log \left( \frac{2^{10} \times H \times W}{\sum_{k=1}^{n} z_k D_k^2 + \sum_{k=1}^{n} \beta_k D_k^2 + \sum_{k=1}^{n} \gamma_k T_k^2} \right),
\]

(20)

where \( n \) means the number of blocks, \( z_k, \beta_k, \) and \( \gamma_k \) counts the number of the pixels for the above three cases in a block, respectively. Assuming \( Q_k = L_{B_k} + 1 \) and \( L_{B_k} = T \), we can deduce

\[
\text{PSNR} \geq 10 \log \left( \frac{2^{10} \times H \times W}{\sum_{k=1}^{n} z_k T_k^2 + \sum_{k=1}^{n} \beta_k (2T+1)^2 + \sum_{k=1}^{n} \gamma_k T_k^2} \right),
\]

(21)

from which it can be seen that PSNR is decreased with the increase of \( T \).

Fig. 5 illustrates the average PSNRs for different kinds of images with different thresholds. In particular, if \( n \) equals the total number of the blocks, and \( \beta_k = \gamma_k = 0 \), Eq. (21) can be further represented as PSNR \( \geq 10 \log (255^2 / T^2) \), which is plotted by the blue solid line with stars in Fig. 5. On the average, the natural and SAR images satisfy this relationship because most of the pixels do not need to be adjusted before embedding. It can also be seen that the medical images have the lowest PSNRs in comparison with others at the same threshold. This lies in the fact that embedding levels in medical images often equal \( T \) due to their high JND thresholds. Moreover, with the increase of \( T \), the number of the unreliable blocks in medical images become ever bigger, i.e., \( \beta_k \) and \( \gamma_k \) in Eq. (20) get increasingly larger, and the PSNRs of medical images therefore drop quickly.

Figs. 6–8 report the robustness of CAR against JPEG and JPEG2000 compression, and AGN with different thresholds. Here, we define the surviving level as a metric to show how much the robustness is, which is expressed by quality factor, bit rate and variance for three attacks, respectively. The surviving level means the embedded watermarks can be extracted correctly when the strength of the attacks is lower than the surviving level. As shown, the larger the threshold is, the stronger the robustness against JPEG compression of natural and medical images is. By contrast, SAR images have the same robustness for different thresholds, i.e., the surviving level is 70. For JPEG2000 compression, medical images achieve the best robustness, up to 0.2 bpp when the threshold is greater than 7, while the other images remain the same, i.e., 0.6 bpp for natural images and 1.4 bpp for SAR images. Moreover, the robustness of SAR images is weaker than that of others. Different from the above attacks, a strong and stable robustness against AGN is obtained for three kinds
of images, e.g., the variance is as high as 0.01 for natural and medical images even when the threshold is equal to 6.

Based on the above analysis, it is essential to select a suitable threshold to balance invisibility and robustness.

4.3. Comparison results

In this sub-section, we first evaluated the proposed CAR in comparison with the HR-based methods in terms of visual quality and reversibility on the assumption that the watermarked images do not suffer from any attacks. Then, we conducted extensive experiments to demonstrate the superiority of the proposed CAR to three typical RLDE methods on the test dataset, wherein the comprehensive performance including capacity, invisibility and robustness is considered when the watermarked images are attacked by the aforementioned attacks. In the experiments, the block size is $8 \times 8$.

4.3.1. "Salt-and-pepper" noise and reversibility

As known, the modulo-256 operation is employed in HR-based methods to prevent the overflow and underflow of pixels. Unfortunately, because the white pixels may be flipped to black ones or black ones to white after modulo operation, HR suffers from severe "salt-and-pepper" noise in the watermarked images. In the proposed CAR, we focus on the flipping of pixels and adopt the PA mechanism to handle this problem. In this experiment, a subjective evaluation is directly given by human observers, since it is probably the best way to assess the visual quality of the watermarked images. Thereafter, PSNR is utilized to offer an objective comparison. Fig. 9 shows the experimental results.
of three example images including Toucan, CT, and S1. Based on
this figure, we have the following observations: (1) the water-
marked images in the second column are seriously degraded
by the “salt-and-pepper” noise due to the flipping of pixels
while the images in the third one are free from such degradation
and almost identical to the host ones from the point of view
of human observers; (2) based on the objective results, it
can be seen that CAR achieves higher PSNRs than HR-based
methods by more than 20 dB. Therefore, we can conclude that
CAR successfully eliminates the “salt-and-pepper” noise, and thus
comfortably outperforms the HR-based methods in terms of
invisibility.

Besides the “salt-and-pepper” noise, the flipped pixels may also
lead to the failure of the recovery of the watermarks and host
images. We term this phenomenon as unreliable reversibility. Let
\( \mathbf{v}_M^A \) and \( \mathbf{v}_M^B \) be the centroid vectors of \( Z_A \) and \( Z_B \) in a block before
embedding while \( \mathbf{v}_{Mw}^A \) and \( \mathbf{v}_{Mw}^B \) be the ones after embedding,
then we make an analysis of the reason why this happens. Fig. 10
illustrates the examples of embedding in regular and flipped cases,
in which the white arrows denote \( \mathbf{v}_M^A \) and \( \mathbf{v}_M^B \) and the black
ones denote \( \mathbf{v}_{Mw}^A \) and \( \mathbf{v}_{Mw}^B \). As shown in Fig. 10(a), \( \mathbf{v}_M^A \) and \( \mathbf{v}_M^B \)
need to be rotated and swapped after embedding so that one can
extract the hidden watermark bit by judging the sign of the
smallest angle between them. However, this condition is violated

Fig. 10. Examples of embedding in (a) regular case, i.e., no flipping of pixels, and (b) flipped case.

Fig. 11. Examples of the host images and recovered ones: (a) host images; (b) recovered images obtained by HR; (c) recovered images obtained by CAR. In (b), the red
squares enclosed by the yellow circles denote the unrecovered blocks. (For interpretation of the references to color in this figure legend, the reader is referred to the web
version of this article.)
when the flipping of pixels occurs, as shown in Fig. 10(b). In such a case, the rotations imposed on the original centroid vectors, $\mathbf{v}_{MA}$ and $\mathbf{v}_{MB}$, are far from being sufficient to swap them, resulting in the errors of watermark extraction. To further validate the unstable reversibility of HR-based methods, Fig. 11 shows some examples of the recovered images to make a comparison in reversibility. The unrecovered blocks in Fig. 11(b) show this drawback, and thus demonstrate the superiority of the proposed CAR method.

### 4.3.2. Comprehensive performance

In practical scenarios, the watermarked images may be degraded to some extent by the aforementioned attacks. In this sub-section, we will investigate the comprehensive performance of the proposed CAR and compare it with other three methods: HR [9], HDC (1) [20], and HDC (2) [26] in this case. It should be noted that the compared methods were conducted on the same experimental setup as CAR to provide a fair comparison. To be specific, the thresholds of CAR as well as HDC (1) and the embedding level of HR are 8, and the shifting quantity of HDC (2) is adaptive to images according to [26]. Additionally, BCH(15, 11, 1) is used in HDC (1) and HDC (2).

Tables 5–7 report the experimental results under the three kinds of attacks, in which $C$ (bpp), $P$ (dB) and $R$ are used to denote the capacity, PSNR and robustness, respectively. To better understand the comparison of robustness, we need to emphasize that the smaller surviving level means stronger robustness for JPEG and JPEG2000 compression but weaker robustness for AGN due to their different units. Based on the experiments, we draw the following conclusions:

1. CAR achieves higher invisibility than HR by more than 10 dB, and even more than 20 dB for medical images, with the same capacity and close robustness, which lies in the fact that PA mechanism adopted in CAR avoids the “salt-and-pepper” noise in the watermarked images; in particular, both invisibility and robustness against JPEG2000 compression for medical images are much higher than that of HR.
2. Both invisibility and capacity of CAR are superior to that of HDC (1) except that the invisibility of medical images is somewhat low, and the robustness against AGN of CAR is up to 0.01, stronger than that of HDC (1).
3. Compared with HDC (2), CAR has the distinct advantages of high capacity and strong robustness and its invisibility, although inferior to that of HDC (2), is enough for practical applications, i.e., human observers cannot distinguish the watermarked images from the host ones.

In a word, the proposed CAR successfully eliminates the “salt-and-pepper” noise in watermarked images and thus improves the invisibility of HR-based methods. Compared with the typical RLDE methods, it achieves better comprehensive performance in terms of capacity, invisibility and robustness.

### 5. Conclusion

In this paper, we have studied the problem of “salt-and-pepper” noise in De Vleeschouwer et al.’s histogram rotation model and developed a robust lossless data embedding (RLDE) method, termed as the content-adaptive reliable RLDE or CAR for short. It changes the grayscale values of the pixels likely to overflow and underflow before embedding by using the pixel adjustment mechanism. Therefore, CAR is free of “salt-and-pepper” noise, greatly improves invisibility and thus constructs a reliable RLDE. To further balance the invisibility and robustness in CAR, we have employed the luminance masking to compute the just noticeable distortion thresholds of local blocks in a host image, which together with the threshold strategy determines different watermark strengths during the embedding procedure, building a content-adaptive reliable embedding. Extensive experimental results on natural, medical and synthetic aperture radar (SAR) images have demonstrated the superiority of the proposed CAR method in comparison with typical RLDE methods.

### Table 5
Comparison of performance under JPEG compression.

<table>
<thead>
<tr>
<th>Image</th>
<th>HR</th>
<th>HDC (1)</th>
<th>HDC (2)</th>
<th>CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C$ (bpp)</td>
<td>$P$ (dB)</td>
<td>$R$</td>
<td>$C$ (bpp)</td>
</tr>
<tr>
<td>Natural</td>
<td>0.015</td>
<td>22.67</td>
<td>50</td>
<td>0.011</td>
</tr>
<tr>
<td>Medical</td>
<td>0.015</td>
<td>8.92</td>
<td>40</td>
<td>0.011</td>
</tr>
<tr>
<td>SAR</td>
<td>0.015</td>
<td>21.35</td>
<td>30</td>
<td>0.011</td>
</tr>
</tbody>
</table>

### Table 6
Comparison of performance under JPEG2000 compression.

<table>
<thead>
<tr>
<th>Image</th>
<th>HR</th>
<th>HDC (1)</th>
<th>HDC (2)</th>
<th>CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C$ (bpp)</td>
<td>$P$ (dB)</td>
<td>$R$</td>
<td>$C$ (bpp)</td>
</tr>
<tr>
<td>Natural</td>
<td>0.015</td>
<td>22.67</td>
<td>0.6</td>
<td>0.011</td>
</tr>
<tr>
<td>Medical</td>
<td>0.015</td>
<td>8.92</td>
<td>0.8</td>
<td>0.011</td>
</tr>
<tr>
<td>SAR</td>
<td>0.015</td>
<td>21.33</td>
<td>0.8</td>
<td>0.011</td>
</tr>
</tbody>
</table>

### Table 7
Comparison of performance under AGN.

<table>
<thead>
<tr>
<th>Image</th>
<th>HR</th>
<th>HDC (1)</th>
<th>HDC (2)</th>
<th>CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C$ (bpp)</td>
<td>$P$ (dB)</td>
<td>$R$</td>
<td>$C$ (bpp)</td>
</tr>
<tr>
<td>Natural</td>
<td>0.015</td>
<td>22.67</td>
<td>0.01</td>
<td>0.011</td>
</tr>
<tr>
<td>Medical</td>
<td>0.015</td>
<td>8.92</td>
<td>0.01</td>
<td>0.011</td>
</tr>
<tr>
<td>SAR</td>
<td>0.015</td>
<td>21.33</td>
<td>0.01</td>
<td>0.011</td>
</tr>
</tbody>
</table>
In the future, we will exploit the selection of the threshold, because it helps to find more optimal watermark strengths. In addition, we plan to draw on the successful experiences of other techniques [4], e.g. moment [10], feature points [11,15,16] and k-means clustering [23], to design a novel RLDE method and to further improve the robustness.

Acknowledgements
The authors would like to thank the editors and anonymous reviewers for their helpful comments and suggestions. This work was supported by the National Basic Research Program of China (973 Program) under Grant 2011CB707104, by the National Natural Science Foundation of China under Grants 61172143, 61072093, 61101250, and 60803150, by the National Science Basic Research Plan in Shaanxi Province of China under Grants 2009JM8004, 2010JQ8026, and 2011JQ8008, and by the Fundamental Research Funds for the Central Universities under Grant K50511030005.

References


[29] Lingling An received the B.S. and M.S. degrees in computer science and technology from Xidian University, Xi'an, Shaanxi, China, in 2002 and 2005, respectively. She is currently pursuing her Ph.D. degree in intelligent information processing from the School of Electronic Engineering, Xidian University. Her current research interests include lossless data hiding and digital watermarking.

Xinbo Gao (M’02–SM’07) received the B.S., M.S., and Ph.D. degrees in signal and information processing from Xidian University, Xi’an, Shaanxi, China, in 1994, 1997 and 1999, respectively. From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, Chinese University of Hong Kong, Shatin, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. Currently, he is a Professor of pattern recognition and intelligent system and the Director of the VIPS Laboratory, Xidian University. His current research interests include computational intelligence, machine learning, computer vision, pattern recognition and wireless communications. In these areas, he has published 4 books and around 100 technical articles in refereed journals and proceedings including IEEE T-IP, T-CSVT, T-NN, T-SMC, Pattern Recognition, and so on. Dr. Gao is on the editorial boards of several journals, including EURASIP Signal Processing (Elsevier), and Neurocomputing (Elsevier). He served as the general chair/co-chair or program committee chair/co-chair or a PC member for around 30 major international conferences.

Yuan Yuan is a researcher (professor) with Chinese Academy of Sciences, and her main research interests include Visual Information Processing and Image/Video Content Analysis.

Dacheng Tao (M’07) received the B.Eng. degree from the University of Science and Technology of China (USTC), Heifei, Anhui, China, the M.Phil. degree from the Chinese University of Hong Kong (CUHK), Shatin, Hong Kong, and the Ph.D. degree from the University of London, London, U.K. Currently, he is a Professor with the Centre for Quantum Computation and Information Systems (CQI) and the Faculty of Engineering and Information Technology in the University of Technology, Sydney, Australia. He mainly applies statistics and mathematics for data analysis problems in data mining, computer vision, machine learning, multimedia, and video surveillance. He has authored and co-authored more than 100 scientific articles at top venues including IEEE T-PAMI, T-KDE, T-IP, NIPS, AISTATS, AAAI, CVPR, ECCV, ICDM; ACM T-KDD, and KDD, with best paper awards. Dr. Tao serves as an Associate Editor of IEEE Transactions on Knowledge and Data Engineering (T-KDE) and the Official Journal of the International Association for Statistical Computing—Computational Statistics & Data Analysis (Elsevier).
Cheng Deng received the B.Sc., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi’an, China. Currently, he is an associate professor with the School of Electronic Engineering at Xidian University. His research interests include information hiding, video copy detection and image quality assessment.

Feng Ji is a Ph.D. student with the School of Electronic Engineering at Xidian University. His research interests are information hiding and reversible watermarking.