Data Hiding in Perceptually Masked OpenEXR Image

KaiLin Chia and KokSheik Wong
School of Information Technology
Monash University Malaysia
Bandar Sunway, Malaysia
{kai.chia, wong.koksheik}@monash.edu

Jean-Luc Dugelay
Department of Digital Security
EURECOM
Biot, France
jean-luc.dugelay@eurecom.fr

Abstract—High dynamic range (HDR) imaging is able to capture and display significantly more colors when compared to the legacy imaging technology. Recently, HDR environment is gaining popularity and becoming common in our daily life, which resulted in the growing number of HDR images. In this work, a technique is put forward to first perceptually mask a HDR image, and then hide data into the masked HDR image. Specifically, pixels in the OpenEXR file, which are stored in half floating point precision format, are manipulated. First, a predictor is utilized to predict pixel values, where well predicted pixel locations are flagged and utilized as the venues to hide data. On the other hand, the ill predicted pixel locations are flagged as unusable. Next, each pixel is divided into segments of 5 bits, and the segments are XOR-ed and permuted to mask the perceptual semantic of the HDR image. Data hiding then takes place at locations flagged as usable, which further distorts the quality of the image. The proposed method is both reversible and separable. The basic performance of the proposed joint technique is evaluated by using 6 HDR images. In the best case scenario, 91% of pixels in the image can be utilized for data hiding purpose while the perceptual semantic of the original image is completely masked.

Index Terms—HDR, OpenEXR, perceptual masking, data insertion

I. INTRODUCTION

Perceptual masking aims to mask the semantic of a content, and it has been heavily relied on for the purpose of secure communication as well as privacy preservation. Since contents are communicated, shared, and stored online these days, digital content encryption, particularly image, audio and video, have received much attention from the research community [1]. On the other hand, data hiding is the act of inserting data into a digital content to serve specific purpose [2]. Some of the typical applications of data hiding include watermarking, steganography, tamper detection and hyperlinking. Data hiding can be applied to any digital content, including traditional contents such as audio, image, video and text document, or un-traditional contents such as IP packets [3] or laser beam [4].

Among various types of digital content, image remains to be a popular one because it can be easily captured, edited, and shared. With the advancement of smart device, even an entry level model can produce high dynamic range (HDR) image, which is able to present a greater range of luminance levels. A cost effective way to obtain a HDR image is to capture several images, each of different level of exposure (viz., underexposed, normal and overexposed) of the same scene [5]. Details captured in each image are then combined to render the scene more realistically. The higher dynamic range enables digital image to store more information about the scene, which is beneficial for applications such as medical imaging and remote sensing because the increase in stored information can provide more precise analysis. To accommodate the natural of HDR, various file formats are put forward, including JPEG2000, JPEG-XR, JPEG-XT [6], OpenEXR [7], Radiance RG BE [8] and LogLUV TIFF [9].

With the emergence of HDR image, there is an increasing need to have proper techniques for managing HDR image and preserving privacy. In that regard, some perceptual encryption techniques are proposed to mask HDR image. For example, Lin et al. [10] proposed an encryption technique for RGBE image by using elementary cellular automata (ECA) function. Yan et al. [11] proposed a format-compliant technique to encrypt LogLUV TIFF image format. The LogLUV pixel values are extracted from the TIFF file and XOR-ed with a stream cipher prior to recompression into a TIFF file. Ting et al. [12] proposed an encryption technique for JPEG XT compressed image. Specifically, DC and AC coefficients in the base and residual layers are shuffled in multiple rounds to completely mask the perceptual semantic of the image. On the other hand, researcher also invented new techniques to hide data into HDR image. Fujiyoshi et al. [13] put forward a blind and reversible method based on properties of floating point values observed in HDR image. The method is inspired by the histogram shifting method [14] designed for traditional integer-based image. Specifically, histogram of an HDR image is found to be sparse and has many empty bins. This characteristic is exploited as venues for reversible data hiding. Chang et al. [15] designed a method for Radiance RGBE [9] file format. A group of homogeneous representations for each pixel is generated by modifying the quadruplet \( \{R, G, B, E\} \), where \( R \), \( G \), \( B \), and \( E \) represent the red, green blue and exponent channels, respectively. Basically, a pixel value can be encoded in different representations.
where each representation can be decoded to the same value. Each representation is associated with \( m = \lfloor \log_2(n) \rfloor \) bits, where \( n \) is the number of homogeneous representations for a color. Then, the representation which is associated to the \( m \)-bit data to be hidden is selected as the new representation.

Often, both perceptual masking and data hiding are jointly utilized. For example, in the hospital setting, a nurse may need to extract patients’ identification information from perceptually masked (or encrypted) CAT scan images. To protect privacy of the patients, medical images are masked, but for management purpose, patients’ information is hidden so that it can be extracted without decrypting the images. Similarly, in cloud storage, user’s contents are encrypted while information about the contents is embedded for administrative purpose. Therefore, the capability to hide data into perceptually masked HDR image is desired. Some conventional joint techniques include [16]–[18], and to the best of our knowledge, none has been proposed for HDR image.

Nonetheless, in [16], the image pixels are encrypted by using stream cipher, then additional data can be embedded by modifying blocks of the encrypted image. Each block is further divided into two sets \( \{S_0, S_1\} \) and the block can be used to hide one bit by flipping the LSB in \( S_0 \) or \( S_1 \), depending on the bit to be hidden. One disadvantage of this method is that the hidden data cannot be extracted before decryption. Therefore, in [17], Zhang improves the method by making it separable, where data can be extracted without viewing the original image content. In [18], to achieve separability, a specific stream encryption algorithm which can preserve correlation between neighboring pixels is proposed. Specifically, the stream encryption algorithm is followed by a permutation (applied to image blocks) to completely mask the image to ensure higher security while preserving the correlation between pixels in the image blocks. After encryption, any previously proposed DHS (difference histogram shifting) and PEHS (prediction-error histogram shifting) based RDH schemes can be deployed to hide additional information.

In this work, a technique is put forward to first perceptually mask a HDR image, then hide data into the perceptually masked image. Our work is designed for the OpenEXR format, which stores floating point values instead of integer values. The proposed method is reversible, where the original image can be perfectly restored, and it is also separable where the hidden data can be extracted without restoring (i.e., unmasking) the image. Experiments are conducted to evaluate the basic performance of the proposed joint technique for HDR image.

II. PRELIMINARIES

A. OpenEXR File Format

OpenEXR is an open source HDR image file format developed by Industrial Light & Magic [7]. It supports higher dynamic range and color precision when compared to the traditional 8-bit file format. OpenEXR file is capable of storing pixels in 16- or 32-bit floating-point data type, and the pixels can be compressed in either lossless or lossy mode. OpenEXR image can have a number of user-defined channels, each with a different data type. The most common ones are the \( R, G \) and \( B \) channels [7].

Specifically, the half precision format is a simplified version of the IEEE-754 floating-point specification [19], which is widely known for its single and double precision floating point formats. Half precision format is made up of 1 sign bit (denoted by \( S \)), 5 exponent bits (denoted by \( E \)), and 10 mantissa bits (denoted by \( M \)). Figure 1 further details the arrangement of each group of bits. Similar to single and double precision formats, the half precision format supports denormalized numbers, positive and negative infinities, and NaN (i.e., not a number). The representable magnitude ranges from \( 2^{-24} \approx 5.96 \times 10^{-8} \) to \( (2^{10}) \times 215 = 65504 \). Table I summarizes these representations for half precision format based on \( E \) and \( M \).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E = 11111_2 )</td>
<td>NaN</td>
</tr>
<tr>
<td>( E = 11111_2 )</td>
<td>( \pm \text{Inf} )</td>
</tr>
<tr>
<td>( E = 0 ) and ( M &gt; 0 )</td>
<td>Subnormal Number</td>
</tr>
<tr>
<td>others</td>
<td>Normalized Number</td>
</tr>
</tbody>
</table>

III. PROPOSED METHOD

Since the exponent part of a pixel determines its coarse value, the exponents of a natural image are highly correlated, particularly for a group of pixels in the same region. This characteristic can be exploited for reversible data hiding, where the exponent values can be replaced by data bits but at the same time they can be completely recovered later through prediction. In addition, the altered exponent values can result in drastic change to the pixel values, causing severe quality degradation. Figure 2 shows the flow of processes in achieving perceptual masking and data hiding.

A. Preprocessing

The exponent part of all pixels is first predicted by using a pixel predictor. The purpose of this step is to identify well predictable locations, so that they can be replaced by the data to be hidden. In this work, we consider the Median Edge Detection (MED) predictor in JPEG-LS [20]. MED is a simple predictor which relies on the direction of the edges, i.e., horizontal, vertical, or no edges at all. The operation is formally described by Eq. (1), where \( x \) is the value to be predicted, \( a, b \) and \( c \) are the left, top and top-left pixels of \( x \),
The prediction error \( e = E_o - E_p \) is computed, where \( E_p \) and \( E_o \) are the predicted and original exponent values, respectively. To flag the usability of each exponent for data hiding purpose, the value \( e \) is analyzed. Specifically, if \( e = 0 \) (viz., the prediction is perfect), the least significant bit (LSB) of \( E_o \) is modified to ‘0’. Otherwise, the LSB of \( E_o \) is set to ‘1’. The original LSB of an ill predicted \( E_o \) is stored in a dynamic array \( C \). As for well predicted \( E_o \), its LSB is not stored because the entire \( E_o \) can be predicted perfectly. Array \( C \) is important to achieve reversibility. Let \( E'_o \) denote the modified exponent value, where \( |E_o - E'_o| \leq 1 \). The LSB of \( E'_o \) then guides the data hiding process, where ‘0’ denotes that data can be hidden (or from the perspective of the receiver - data is hidden), while ‘1’ means that no data is hidden.

### B. Masking The Image

First, the 10-bit mantissa part \( M \) is divided and treated as two independent blocks each with 5 bits, namely, \( M_1 \) and \( M_2 \). The subblocks \( M_1 \) and \( M_2 \) each assumes a size of 5 bits, which is the same as the length of \( E_o \) to ease the masking operations. Therefore, each pixel will assume the form of \([S, E'_o, M_1, M_2]\), but for the purpose for this work, \( S \) is ignored. Second, \( M_1 \) and \( M_2 \) are extracted from all pixels, and shuffled by using a key \( \kappa_1 \). The shuffled \( M_1 \) and \( M_2 \) form new mantissas (see Fig. 4). The potential outcomes for each pixel include:

\[
\begin{align*}
[E'_o(i), M_1(j), M_2(k)], \\
[E'_o(i), M_2(j), M_1(k)], \\
[E'_o(i), M_1(j), M_1(k)], \\
[E'_o(i), M_2(j), M_2(k)], & \text{ if } j \neq k; \quad \text{or}, \\
[E'_o(i), M_2(j), M_2(k)], & \text{ if } j \neq k,
\end{align*}
\]

where \( M_1(j) \) denotes the \( M_1 \) block from the \( j \)-th pixel, and \( M_2(k) \) as well as \( E'_o(i) \) are defined in same manners. The purpose of Stage II is to mask the details or textures (particularly recurring patterns) captured by the mantissa part, although mantissa can only cause small variation of value in comparison to the exponent part.

Next, the 4 most significant bits (4MSB) of the exponents \( E'_o \) for each pixel as well as its newly formed mantissa are XOR-ed with a random sequence of 0’s and 1’s, which is generated with the key \( \kappa_2 \) (Stage III). Figure 5 illustrates an example, where \( E''_o \) and \( M'' \) refer to the modified exponent and mantissa parts, respectively. Note the LSB of \( E'_o \) does not undergo the XOR step to maintain the flag, i.e., \( \text{LSB}(E'_o) = \text{LSB}(E''_o) \). This XOR step is important to break the high correlation between the pixels, particularly the ill predicted pixels, as they will not be involve the data hiding process in Stage V (see Section III-C for detailed explanation). The newly formed pixels are then shuffled by using the key \( \kappa_3 \) (Stage IV) to further increase the overall entropy of the image pixels.

It is noteworthy that Stages II, III and IV are independent to each other. Therefore, they can be implemented without following any particular order. In addition, a master key \( \kappa_m \) for content masking can be considered to generate the keys \( \kappa_1, \kappa_2 \) and \( \kappa_3 \).

**Figure 2:** Overall process of proposed method.

**Figure 3:** Layout of MED predictor.

\[
x' = \begin{cases} 
  \min(a, b), & \text{if } c \geq \max(a, b); \\
  \max(a, b), & \text{if } c \leq \min(a, b); \\
  a + b - c, & \text{otherwise.}
\end{cases}
\]  

\[ (1) \]

MED predictor predicts all values in one pass following the raster scan order, where pixels at the boarder are utilized as the initial reference pixels.

respectively, as shown in Figure 3. The predicted value of \( x \) is computed as follows:

The prediction error \( e = E_o - E_p \) is computed, where \( E_p \) and \( E_o \) are the predicted and original exponent values, respectively. To flag the usability of each exponent for data hiding purpose, the value \( e \) is analyzed. Specifically, if \( e = 0 \) (viz., the prediction is perfect), the least significant bit (LSB) of \( E_o \) is modified to ‘0’. Otherwise, the LSB of \( E_o \) is set to ‘1’. The original LSB of an ill predicted \( E_o \) is stored in a dynamic array \( C \). As for well predicted \( E_o \), its LSB is not stored because the entire \( E_o \) can be predicted perfectly. Array \( C \) is important to achieve reversibility. Let \( E'_o \) denote the modified exponent value, where \( |E_o - E'_o| \leq 1 \). The LSB of \( E'_o \) then guides the data hiding process, where ‘0’ denotes that data can be hidden (or from the perspective of the receiver - data is hidden), while ‘1’ means that no data is hidden.

### B. Masking The Image

First, the 10-bit mantissa part \( M \) is divided and treated as two independent blocks each with 5 bits, namely, \( M_1 \) and \( M_2 \). The subblocks \( M_1 \) and \( M_2 \) each assumes a size of 5 bits, which is the same as the length of \( E_o \) to ease the masking operations. Therefore, each pixel will assume the form of \([S, E'_o, M_1, M_2]\), but for the purpose for this work, \( S \) is ignored. Second, \( M_1 \) and \( M_2 \) are extracted from all pixels, and shuffled by using a key \( \kappa_1 \). The shuffled \( M_1 \) and \( M_2 \) form new mantissas (see Fig. 4). The potential outcomes for each pixel include:

\[
\begin{align*}
[E'_o(i), M_1(j), M_2(k)], \\
[E'_o(i), M_2(j), M_1(k)], \\
[E'_o(i), M_1(j), M_1(k)], \\
[E'_o(i), M_2(j), M_2(k)], & \text{ if } j \neq k; \quad \text{or}, \\
[E'_o(i), M_2(j), M_2(k)], & \text{ if } j \neq k,
\end{align*}
\]

where \( M_1(j) \) denotes the \( M_1 \) block from the \( j \)-th pixel, and \( M_2(k) \) as well as \( E'_o(i) \) are defined in same manners. The purpose of Stage II is to mask the details or textures (particularly recurring patterns) captured by the mantissa part, although mantissa can only cause small variation of value in comparison to the exponent part.

Next, the 4 most significant bits (4MSB) of the exponents \( E'_o \) for each pixel as well as its newly formed mantissa are XOR-ed with a random sequence of 0’s and 1’s, which is generated with the key \( \kappa_2 \) (Stage III). Figure 5 illustrates an example, where \( E''_o \) and \( M'' \) refer to the modified exponent and mantissa parts, respectively. Note the LSB of \( E'_o \) does not undergo the XOR step to maintain the flag, i.e., \( \text{LSB}(E'_o) = \text{LSB}(E''_o) \). This XOR step is important to break the high correlation between the pixels, particularly the ill predicted pixels, as they will not be involve the data hiding process in Stage V (see Section III-C for detailed explanation). The newly formed pixels are then shuffled by using the key \( \kappa_3 \) (Stage IV) to further increase the overall entropy of the image pixels.

It is noteworthy that Stages II, III and IV are independent to each other. Therefore, they can be implemented without following any particular order. In addition, a master key \( \kappa_m \) for content masking can be considered to generate the keys \( \kappa_1, \kappa_2 \) and \( \kappa_3 \).

**Figure 2:** Overall process of proposed method.

**Figure 3:** Layout of MED predictor.

\[
x' = \begin{cases} 
  \min(a, b), & \text{if } c \geq \max(a, b); \\
  \max(a, b), & \text{if } c \leq \min(a, b); \\
  a + b - c, & \text{otherwise.}
\end{cases}
\]  

\[ (1) \]
Figure 4: An example output of mixing $M_1$ and $M_2$ blocks among pixels (Stage II).

Figure 5: Masking pixel using XOR operation (Stage III).

C. Data Hiding

Let $\phi$ be the external data to be hidden, which is usually encrypted with a key $\kappa_4$. For reversibility purpose, the original LSB of the exponent, denoted by $C$ needs to be stored. Therefore, the augmented data $\Phi = [C|\phi]$ is constructed, where $|$ refers to the concatenation operation. $\Phi$ is then processed in segments of 4-bit. Data hiding is carried out by considering pixel locations where $\text{LSB}(E''_o) = 0$ for $E''_o$ being the exponent part of the pixel at Stage IV. Specifically, for each usable pixel, the most significant 4 bits of its exponent are replaced by a 4-bit segment from $\Phi$. Data hiding will not take place if $\text{LSB}(E''_o) = 1$. The process is repeated until all segments of $\Phi$ are inserted or when one runs out of usable pixels. Figure 6 illustrates the data insertion process, where the new exponent value is denoted as $E''_o$. (a) Case 1: $E''_o$ with LSB of ‘0’. (b) Case 2: $E''_o$ with LSB of ‘1’

Figure 6: Hiding data into preprocessed exponent part.

D. Data Extraction and Image Recovery

The data extraction and image recovery steps are exactly the reverse of the encryption and data insertion processes. Specifically, the inserted data, i.e., $\Phi$ is first extracted, and it is processed to obtain $\phi$ and $C$. The intended receiver will be able to decrypt $\phi$ using $\kappa_4$. Note that the data can be extracted without the need to decrypt the image.

To recover the original image, the encryption processes need to be reversed, i.e., inverse shuffling in Stage IV using $\kappa_3$, XOR operation in Stage III using the same binary sequence generated by $\kappa_2$, and inverse shuffling of the $M_1$’s and $M_2$’ blocks in Stage II using $\kappa_1$. Next, the LSB of the exponent part for each pixel is examined to reconstruct the image. Specifically, when $\text{LSB}(E) = ‘1’$, that means 4MSB planes were not replaced, i.e., the original sequence is intact, and the first bit of $C$ is removed and used as the LSB for $E$. On the other hand, if $\text{LSB}(E) = ‘0’$, then 4MSB planes are recovered through prediction. Note that the predict-to-recover process must be carried out in raster scan order again to ensure the bits from $C$ is reinserted at the correct position, as only pixels flagged as unpredictable have their original LSB stored in $C$ (length of array $C \neq$ total pixels in image).

IV. EXPERIMENT

The proposed algorithm is implemented in Matlab R2018a. 6 images from the dataset [21] are selected randomly for experiment purpose. Since the original image in [21] is large in dimension (i.e., larger than 4K), the selected images are down-scaled to 6.25% of the original size. These images are shown in Fig. 7. It is confirmed that the inserted data can be extracted, and the original image can be completely reconstructed from its processed counterpart.

A. Effective Payload

First, the raw capacity achieved for each image using the proposed algorithm is computed as:

$$\frac{\text{number of usable pixels}}{\text{total pixels}} \times 100\%, \quad (3)$$

while the percentage of side information is computed as:

$$\frac{\text{number of unusable pixels}}{\text{number of usable pixels}} \times 100\%. \quad (4)$$

Here, raw capacity refers to the actual number of pixels that are flagged as suitable for hiding data, while side information percentage is the fraction of raw capacity spent on coding the array $C$. Table II records the capacity for each test image. Results suggest that the percentage of usable pixels is high for all HDR images considered, with an average of more than
TABLE II: Hiding capacity for different test images.

<table>
<thead>
<tr>
<th>Image</th>
<th>raw capacity (%)</th>
<th>side info capacity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>507</td>
<td>89.2549</td>
<td>2.9442</td>
</tr>
<tr>
<td>CanadianFalls</td>
<td>74.0750</td>
<td>8.6622</td>
</tr>
<tr>
<td>CemeteryTree(2)</td>
<td>63.5323</td>
<td>14.2582</td>
</tr>
<tr>
<td>Frontier</td>
<td>86.3199</td>
<td>3.8944</td>
</tr>
<tr>
<td>BigFootPass</td>
<td>90.5954</td>
<td>2.5238</td>
</tr>
<tr>
<td>BalancedRock</td>
<td>88.5727</td>
<td>3.1595</td>
</tr>
</tbody>
</table>

70%. However, some images (e.g., CemeteryTree(2) ) yield lower raw capacity because there are less well predictable pixels. In other words, higher number of ill predictable pixels result in relatively larger side information \( C \), and hence less available space to insert data.

B. Image Quality

This section presents the quality of the processed image from Stage V in two scenarios: the first one hides data after the prediction stage by skipping Stage II, III and IV, and; the second one hides data after the encryption stage, viz., stepping through Stage II, III and IV. By visual inspection, output image in the first scenario is already severely distorted. However, some details of the image are still visible (e.g., see outline of wheel at the lower left region in Fig. 8(c)). Nonetheless, the image produced by in the second scenario is completely distorted, i.e., there is no trace of the original perceptual semantic. This observation verifies that although replacing the exponent part with payload bits already result in a highly distorted image, Stage II, III and IV are crucial to further intensify distortion to completely mask the perceptual semantic of an image. Figure 8 shows two representative output images for both scenarios.

To quantify distortion, log-based PSNR [22] is computed for the processed images for both scenarios, and the results are recorded in Table III. The logarithmic function is an approximation on how Human Visual System (HVS) responds to the visible luminance range. The capability to mimic how human eye response to light makes this quality metric more suitable in capturing the quality of HDR image. As expected, the PSNR values are low (i.e., close to 0) and this confirms that the processed images are highly distorted after applying the proposed image masking and data hiding technique.

TABLE III: Quality assessment after processing (dB).

<table>
<thead>
<tr>
<th>Image</th>
<th>Log PSNR Scenario 1</th>
<th>Log PSNR Scenario 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>507</td>
<td>3.3252</td>
<td>2.7679</td>
</tr>
<tr>
<td>CanadianFalls</td>
<td>4.4658</td>
<td>3.0105</td>
</tr>
<tr>
<td>CemeteryTree(2)</td>
<td>4.8320</td>
<td>2.7357</td>
</tr>
<tr>
<td>Frontier</td>
<td>3.3314</td>
<td>2.9717</td>
</tr>
<tr>
<td>BigFootPass</td>
<td>3.9878</td>
<td>3.4865</td>
</tr>
<tr>
<td>BalancedRock</td>
<td>3.4036</td>
<td>2.7252</td>
</tr>
</tbody>
</table>

C. Discussion

Last but not least, the proposed method is suitable to be applied for floating point numbers, specifically half precision format. The conventional methods [16]–[18] are not suitable for floating point numbers because they are designed for integer pixels, which have different intrinsic properties. For example, 3LSB are manipulated to hide data in [16], but the technique is not readily applicable to pixels stored in floating point format. Specifically, the mantissa part of the pixel in half-precision format is rather random and it is challenging to predict them precisely.
V. CONCLUSIONS

In this paper, image masking and data insertion techniques are put forward for HDR image stored in the OpenEXR format. Specifically, the exponent part of each pixel is classified as predictable or not predictable. The predictable pixels are modified to hide data, while side information is stored as part of the payload for perfect image recovery purpose. To further intensify the distortion, mantissa subblocks shuffling, XOR cipher and inter-pixel shuffling are applied to the pixels. Experiments suggest that, on average, the number of usable pixels is greater than 70% of the total number of pixels. Results also suggest that the quality of an image can be completely distorted and perfectly recovered later when required.

As future work, more detailed analysis will be performed to investigate into the security aspect of the proposed method. In addition, the file structure of OpenEXR will be further explored to improve capacity and secrecy.

REFERENCES


