A Visually-Lossless Data Hiding Method Based on Histogram Modification

Masaaki FUJIYOSHI and Hitoshi KIYA
Department of Information and Communication Systems, Tokyo Metropolitan University, Hino, Tokyo 191–0065, Japan
Email: mfujiyoshi@ieee.org, kiya@sd.tmu.ac.jp

***Abstract*—This paper proposes a new method of visually-lossless data hiding. The proposed method firstly applies a nonlinear quantization to an original image accordingly to the size of a payload to be hidden. The method, then, embeds the payload to the quantized image based on a histogram modification-based lossless data hiding technique. The hidden payload can be extracted reversibly and be completely removed from the image to recover the quantized image. Experimental results show the effectiveness of the proposed method.**

I. Introduction

Data hiding technology has been diligently studied, for not only security-related problems [1], [2], in particular, intellectual property rights protection of digital contents [3], but also non security-oriented [1], [4] such as broadcast monitoring [5]. A data hiding technique embeds data referred to as a *payload* into a target signal referred to as an *original* signal. It, then, generates a slightly distorted signal that is referred to as a *stego* signal.

Many of data hiding techniques extract the hidden payload but leave a stego signal as it is, i.e., *lossy*. Later, *lossless* data hiding in which the original signal is completely recovered as well as the hidden payload is extracted from a stego signal has been proposed [6], [7] for medical and military applications. As in compression, *visually-lossless* concept has been introduced to data hiding [8], [9], besides methods for visually-lossless compression have been developed [10].

Visually-lossless data hiding techniques can be classified to two groups; lossy based [5], [8] and lossless based [9]. The former embeds a payload to an original signal with the lossy data hiding technique, but it generates a stego signal with a user defined image quality [5], [8]. The latter firstly modifies an original signal with a lossy manner as a preprocessing step, then it hides a payload to the preprocessed image with a lossless data hiding technique [9].

This paper proposes a new method for the latter group. The proposed method applies a nonlinear quantization to an original image, and then a payload is hidden to the quantized image based on a histogram modification-based lossless data hiding (HM-LDH) technique [7]. The proposed method not only extracts the hidden payload but also removes the payload from a stego image to recover the quantized image. This method may suit labeling [11] and steganography [12].

II. Preliminary

This section briefly describes the conventional HM-LDH method [7] and its limitation.

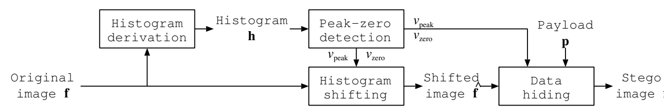


Fig. 1. Conventional histogram modification-based lossless data hiding technique [7].

A. Histogram Modification-Based Lossless Data Hiding

It mentions here the essential of the conventional method [7] which will be partially employed in the proposed method. It is assumed here that L-bits string represented by $\mathbf{p} = \{p(l)\}$, where $p(l) \in \{0,1\}$ and $l = 0,1,\ldots,L-1$, is hidden to $X \times Y$-sized Q-bits quantized grayscale image $\mathbf{f} = \{f(x,y)\}$ where $f(x,y) \in \left\{0,1,\ldots,2^Q-1\right\}$, $x = 0,1,\ldots,X-1$, and $y = 0,1,\ldots,Y-1$.

As shown in Fig. 1, this method firstly derives histogram

$$\mathbf{h} = \{h(v)\} \tag{1}$$

for original image $\mathbf{f}$, where $h(v)$ represents the frequency (number of pixels) for pixel value v in $\mathbf{f}$ and $v = 0,1,\ldots,2^Q-1$. The method, then, finds pixel value v_{peak} corresponding to the maximum frequency $\max_v h(v)$ and pixel value v_{zero} corresponding to zero frequency, i.e.,

$$h\left(v_{\text{peak}}\right) = \max_v h(v), \tag{2}$$

$$h\left(v_{\text{zero}}\right) = 0. \tag{3}$$

For ease of explanation, it is assumed here that

$$L \leq h\left(v_{\text{peak}}\right), \tag{4}$$

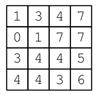
$$v_{\text{zero}} < v_{\text{peak}}. \tag{5}$$

This method prepares a room to hide payload $\mathbf{p}$ in original image $\mathbf{f}$ by subtracting one from all pixels having pixel values between $\left(v_{\text{zero}}+1\right)$ and $\left(v_{\text{peak}}-1\right)$;

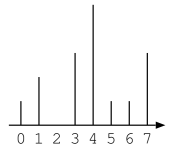
$$\hat{f}(x,y) = \begin{cases} f(x,y)-1, & v_{\text{zero}} < f(x,y) < v_{\text{peak}} \\ f(x,y), & \text{others} \end{cases}, \tag{6}$$

where $\hat{f}(x,y)$'s are pixels in histogram-shifted image $\hat{\mathbf{f}} = \left\{\hat{f}(x,y)\right\}$. The method, now, hides $\mathbf{p}$ to $\hat{\mathbf{f}}$; According to data bit $p(l)$, the pixel having pixel value v_{peak} is decreased by one or is left as is;

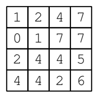
$$\tilde{f}(x,y) = \begin{cases} \hat{f}(x,y)-1, & \hat{f}(x,y) = v_{\text{peak}} \text{ and } p(l) = 0 \\ \hat{f}(x,y), & \text{others} \end{cases}, \tag{7}$$



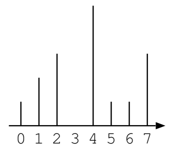
(a) Original image **f**.



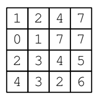
(b) Original histogram **h**.



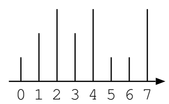
(c) Shifted image $\hat{\mathbf{f}}$.



(d) Shifted histogram $\hat{\mathbf{h}}$.



(e) Stego image $\tilde{\mathbf{f}}$.



(f) Stego histogram $\tilde{\mathbf{h}}$.

Fig. 2. An example of data hiding by the conventional histogram modification-based lossless data hiding method [7]. Five bits data '10110' is hidden to a 3-bits quantized image.

where the stego image is $\tilde{\mathbf{f}} = \left\{\tilde{f}(x,y)\right\}$.

Figure 2 shows an example of this method in which $v_{\text{peak}} = 4$, $h\left(v_{\text{peak}}\right) = 5$, and $p_{\text{zero}} = 2$. In this example, 5-bits binary data '10110' is hidden to a 3-bits quantized original image, i.e., $L = 5$ and $Q = 3$

It is noted that a HM-LDH method generally has to memorize parameter v_{peak} to extract and remove hidden payload **p** from stego image $\tilde{\mathbf{f}}$. After removing **p**, $\tilde{\mathbf{f}}$ becomes histogram-shifted image $\hat{\mathbf{f}}$. Parameter v_{zero}, now, is required to recover original image **f** from $\hat{\mathbf{f}}$.

B. Limitations of Technique

This section mentions two limitations of the conventional HM-LDH method [7] described in the previous section, namely, capacity limitation and no pixel value with zero frequency.

1) Capacity Limitation: As Eq. (7) indicates, the capacity for lossless data hiding of original image **f** in this method is $h\left(v_{\text{peak}}\right)$. That is, Eq. (4) should be satisfied for hiding L-bits payload **p** to **f**. The method can increases the capacity of **f** by using multiple peak-zero pairs [7], it is complicated and multiple parameter pairs should be memorized.

2) No Pixel Value with Zero Frequency: The method needs v_{zero}, a pixel value with zero frequency, in original image **f** as shown in Eq. (3). For **f** without zero frequency pixel value, the value in all pixels having $v_{\min} = \arg\min_v h(v)$ is changed to another to make $v_{\min}$ becoming v_{zero} before histogram shifting [7]. Positions of all $v_{\min}$-valued pixels should be memorized to recover **f**.

In practice, these situations often become problems, even the bypass is given for each problem as mentioned above. The next section proposes a new visually-lossless data hiding method to overcome these problems.

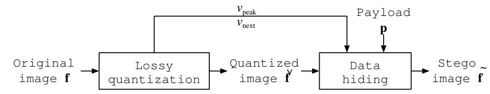


Fig. 3. Proposed method with the introduced quantizer.

III. Proposed Method

This section proposes a new method of visually-lossless data hiding which is shown in Fig. 3. To accept payload **p** overweight for original image **f**, the proposed method quantizes **f** to make a enough room for **p**. By an embedding manner based on the conventional HM-LDH method [7], **p** is hidden to the quantized image.

A. Algorithms

The proposed method consists of three algorithms; lossy quantization, lossless data hiding, and hidden data extraction and quantized image recovery.

1) Lossy Quantization: First of all, the proposed method quantizes original image **f** to make an enough room for hiding L-bits payload **p**.

Step 1. Apply a quantization to **f**. For quantized image $\check{\mathbf{f}} = \left\{\check{f}(x,y)\right\}$ where $\check{f}(x,y) \in \left\{0,1,\ldots,2^Q-1\right\}$, histogram $\check{\mathbf{h}} = \left\{\check{h}(v)\right\}$ should satisfy

$$\check{h}\left(v_{\text{peak}}\right) \geq L, \tag{8}$$

$$\check{h}\left(v_{\text{zero}}\right) = 0. \tag{9}$$

It is assumed again that $v_{\text{zero}} < v_{\text{peak}}$. Though any quantizer can be employed for this algorithm, it is desired that the squared quantization difference (SSD) between **f** and $\check{\mathbf{f}}$ is minimized. One simple quantizer is shown in Sect. III-B.

2) Lossless Data Hiding: Then, the proposed method hides **p** to quantized image $\check{\mathbf{f}}$ based on the conventional HM-LDH method [7].

Step 1. Applying the histogram shifting to $\check{\mathbf{f}}$ to generate histogram-shifted image $\acute{\mathbf{f}} = \{\acute{f}(x,y)\}$ as

$$\acute{f}(x,y) = \begin{cases} \check{f}(x,y) - 1, & v_{\text{zero}} < \check{f}(x,y) < v_{\text{peak}} \\ \check{f}(x,y), & \text{others} \end{cases} \tag{10}$$

Step 2. According to data bit $p(l)$, the pixel having pixel value v_{peak} is decreased by one or is left as is;

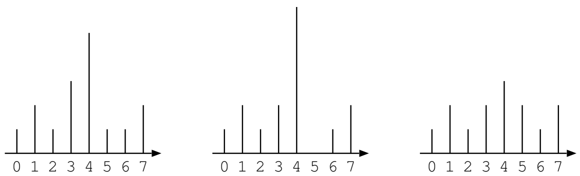
$$\tilde{f}(x,y) = \begin{cases} \acute{f}(x,y) - 1, & \acute{f}(x,y) = v_{\text{peak}} \text{ and } p(l) = 0 \\ \acute{f}(x,y), & \text{others} \end{cases} \tag{11}$$

The quantizer mentioned in Sect. III-B also minimizes the SSD between **f** and $\tilde{\mathbf{f}}$.

3) Hidden Data Extraction and Quantized Image Recovery: From stego image $\tilde{\mathbf{f}}$, hidden payload **p** is extracted and is removed to recover quantized image $\check{\mathbf{f}}$.

Step 1. With memorized parameters v_{peak} and v_{zero}, hidden payload bit $p(l)$ is extracted as

$$p(l) = \begin{cases} 1, & \tilde{f}(x,y) = v_{\text{peak}} \\ 0, & \tilde{f}(x,y) = v_{\text{peak}} - 1 \end{cases}. \tag{12}$$



(a) Original $\mathbf{h}$. (b) Quantized $\check{\mathbf{h}}$. (c) Stego $\tilde{\mathbf{h}}$.

Fig. 4. An example of the histogram modification in the proposed method. 6-bits payload is hidden over pixels having pixel value 4 or 5 ($L=6$, $v_{\text{peak}}=4$, $h\left(v_{\text{peak}}\right)=5$, $\check{h}\left(v_{\text{peak}}\right)=C=7$, $v_{\text{next}}=v_{\text{peak}}+1$).

Step 2. With the memorized parameters, quantized image $\check{\mathbf{f}}$ is recovered as

$$\check{f}(x,y)=\begin{cases}\tilde{f}(x,y)+1, & v_{\text{zero}}\leq\tilde{f}(x,y)\leq v_{\text{peak}}-1\\ \tilde{f}(x,y), & \text{others}\end{cases}. \quad (13)$$

The quantizer shown in the next section makes the proposed method free from memorizing v_{zero}.

B. An Example Quantizer

A simple quantizer is introduced for the proposed method. This quantizer finds a pixel value pair v_{peak} and v_{next} where

$$h\left(v_{\text{peak}}\right)+h\left(v_{\text{next}}\right)=C\geq L, \quad (14)$$

$$\left|v_{\text{peak}}-v_{\text{next}}\right|=1. \quad (15)$$

Even original image $\mathbf{f}$ may have several pairs satisfying Eqs. (14) and (15), the quantizer picks the pair having the maximum $h\left(v_{\text{peak}}\right)$ among the pairs. Then, $\mathbf{f}$ is quantized as

$$\check{f}(x,y)=\begin{cases}v_{\text{peak}}, & f(x,y)=v_{\text{next}}\\ f(x,y), & \text{others}\end{cases}, \quad (16)$$

i.e., $v_{\text{zero}}=v_{\text{next}}$, and histogram shifting in Step 1 of the algorithm in Sect. III-A2 can be avoided, i.e., $\acute{\mathbf{f}}=\check{\mathbf{f}}$.

Figure 4 shows an example of data hiding by the proposed method with this quantizer. In this example, 6-bits payload is hidden to 3-bits quantized image $\mathbf{f}$ with 16 pixels, i.e., $L=6$, $Q=3$, and $XY=16$. As shown in Fig. 4 (a), $\mathbf{f}$ has peak frequency $h\left(v_{\text{peak}}\right)=5$ which is less than L, c.f., Sect. II-B1. In addition, no zero frequency is in $\mathbf{h}$, c.f., Sect. II-B2. This quantizer modifies $\mathbf{f}$ to satisfy Eqs. (8) and (9). To quantized image $\check{\mathbf{f}}$ with $C=7\geq L=6$ (Fig. 4 (b)), 6-bits payload is hidden (Fig. 4 (c)).

C. Features

This section summarizes two main features of the proposed method. That is, visually-lossless data hiding and quantization based on histogram modification.

1) Visually-Lossless Data Hiding: The proposed method introduces the visually-lossless concept into HM-LDH to overcome two problems mentioned in Sect. II-B.

The proposed method accepts a payload overweight for an original image by a simple strategy; Increasing the peak frequency of the histogram with a lossy manner instead of finding, using, and memorizing multiple peak-zero pairs for defending lossless data hiding [7] as mentioned in Sect. II-B1. Since it is not guaranteed that multiple peak-zero pairs always exist in any image, using multiple peak-zero pairs may occurs another problem; no pixel value with zero frequency.

As shown in Sect. IV, a user often faces an image in which no pixel value with zero frequency is. The proposed method overcomes this problem with a simple strategy; changing pixel values of certain pixels to another value with a lossy manner to ensure that a zero frequency pixel value exists instead of memorizing positions of pixels having the minimum frequency [7] as mentioned in Sect. II-B2.

The quantization based on histogram modification described in the next section realizes the above mentioned two simple processes simultaneously.

It is noted that the conventional lossless data hiding-based visually-lossless data hiding method [9] is based on a spread spectrum-based lossless data hiding technique [13]. This method changes pixel values of pixels having upper or lower bound values, i.e., near 0 or 255, to inner values so that a modulo arithmetic does not introduce a noticeable distortion. So, the proposed method differs substantially from this conventional method.

2) Quantization Based on Histogram Modification: A lossy quantization is introduced to the proposed method to modify the histogram of an original image. As Eqs. (3) and (4) indicate, a HM-LDH method generally requires the zero frequency and the peak frequency which is greater than or equal to the size of a payload. Instead of the complicated bypass mentioned in Sect. II-B, a quantizer is introduced to satisfy these two requirements.

Though any quantizer can be employed in the proposed method, it has to take account into the histogram of the quantized image. Well-known Lloyd-Max quantizer [14] minimizes the SSD between original and quantized images, it, however, cannot satisfy Eqs. (3) and (4) automatically, in general. Except for such requirement, an appropriate quantizer can be used to control the trade off between the capacity improvement and image degradation.

The quantizer mentioned in Sect. III-B uses not $v_{\min}$ but v_{next} to satisfy Eq. (3) and simultaneously to minimize the SSD between $\mathbf{f}$ and $\tilde{\mathbf{f}}$ by free from histogram shifting. The SSD between $\mathbf{f}$ and $\check{\mathbf{f}}$ is

$$h\left(v_{\text{next}}\right) \quad (17)$$

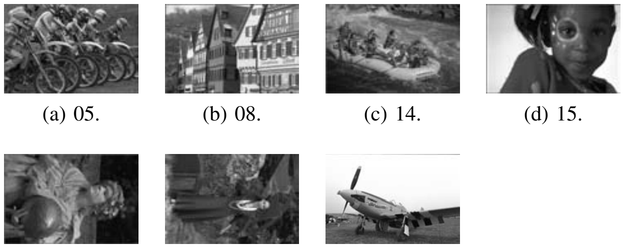
and that between $\mathbf{f}$ and $\tilde{\mathbf{f}}$ is

$$\frac{L}{2C}+\frac{(C-L)h\left(v_{\text{next}}\right)}{C^2}. \quad (18)$$

IV. Experimental Results

Figure 5 shows seven 768×512-sized 8-bits quantized grayscale images from CIPR-RPI [15]. These images do not have zero frequency pixel value.

Figure 6 shows the peak signal-to-noise ratio (PSNR) of quantized images for various payload sizes which the PSNR corresponds to Eq. (17).



(a) 05. (b) 08. (c) 14. (d) 15.

(e) 17. (f) 18. (g) 20.

Fig. 5. Seven images having no pixel value with zero frequency from CIPR-RPI [15].

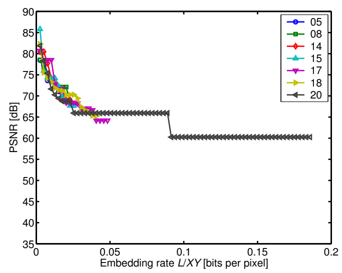


Fig. 6. PSNR versus embedding rate for quantized images.

The PSNR between stego and original images are plotted in Fig. 7. Ten different random payloads which each payload consists of equiprobable zeros and ones are hidden to ten different randomly selected L of C pixels with $v_{\rm peak}$, and the PSNR's are averaged. In the conventional HM-LDH method [7], L takes account into but not include bzip2-compressed pixel positions and pixel value $v_{\rm min}$. With the simple quantizer, the proposed method not only accepts larger payloads but also improves the PSNR's of stego images.

Figure 8 shows results for Fig. 5 (g), and it is confirmed that no noticeable difference is found among images.

V. Conclusions

This paper has proposed a new visually-lossless data hiding method for images. The proposed method hides a payload losslessly to a lossy quantized image, and the hidden payload is not only extracted but also removed from a stego image to recover the quantized image.

Designing a sophisticated quantizer for the proposed method, steganalysis [12] of the proposed method, and applying the usage of multiple successive pixel values having zero frequencies next to the value with the peak frequency [16], [17] to the proposed method are the further works.

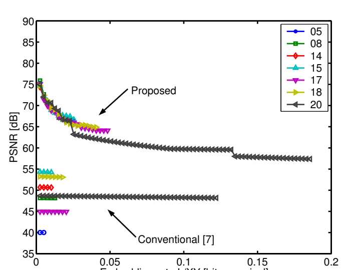


Fig. 7. PSNR versus embedding rate for stego images in the conventional histogram modification-based lossless data hiding [7] and proposed methods.

(a) Original. (b) Quantized (PSNR: 60.26 dB). (c) Stego (PSNR: 57.37 dB).

Fig. 8. Example images for Fig. 5 (g) (Embedding rate: 0.1856 bits per pixel

Acknowledgment

This work has been partly supported by the Grant-in-Aid for Young Scientists (B), No.20336522, from the Ministry of Education, Culture, Sports, Science and Technology of Japan and from the Japan Society for the Promotion of Science.

References

[1] G.C. Langelaar, I. Setyawan, and R.L. Lagendijk, "Watermarking digital image and video data," *IEEE Signal Process. Mag.*, vol.17, no.5, pp.20–46, Sep. 2000.

[2] M. Barni, "What is the future for watermarking? (Part I)," *IEEE Signal Process. Mag.*, vol.20, no.5, pp.55–59, Sep. 2003.

[3] C.-C.J. Kuo, T. Kalker, and W. Zhou, eds., "Digital rights management," *IEEE Signal Process. Mag.*, vol.21, no.2, pp.11–117, Mar. 2004.

[4] M. Barni, "What is the future for watermarking? (Part II)," *IEEE Signal Process. Mag.*, vol.20, no.6, pp.53–59, Nov. 2003.

[5] T. Tachibana, M. Fujiyoshi, and H. Kiya, "A removable watermarking scheme retaining the desired image quality," in *Proc. IEEE ISPACS*, 2003, pp.538–542.

[6] R. Caldelli, F. Filippini, and R. Becarelli, "Reversible watermarking techniques: an overview and a classification," *EURASIP J. Inf. Security*, vol.2010, 2010.

[7] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol.16, pp.354–362, Mar. 2006.

[8] M. Barni, F. Bartolini, V. Cappellini, E. Magli, and G. Olmo, "Near-lossless digital watermarking for copyright protection of remote sensing images," in *Proc. IEEE IGARSS*, vol.3, 2002, pp.1447–1449.

[9] J.H.K. Wu, R.-F. Chang, C.-J. Chen, C.-L. Wang, T.-H. Kuo, W.K. Moon, and D.-R. Chen, "Tamper detection and recovery for medical images using near-lossless information hiding technique," *J. Digital Imaging*, vol.21, pp.59–76, Mar. 2008.

[10] Y.-S. Kim, O-H. Kwon, and R.-H. Park, "Wavelet based watermarking method for digital images using the human visual system," *IEE Electron. Lett.*, vol.35, pp.466–468, Mar. 1999.

[11] Y. Seki, H. Kobayashi, M. Fujiyoshi, and H. Kiya, "Quantization-based image steganography without data hiding position memorization," in *Proc. IEEE ISCAS*, 2005, pp.4987–4990.

[12] J. Fridrich, *Steganography in digital media*, Cambridge University Press, 2010.

[13] C.W. Honsinger, P.W. Jones, M. Rabbani, and J.C. Stoffel, *Lossless recovery of an original image containing embedded data*, U.S. Patent, no.6278791, 2001.

[14] S.P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol.IT-28, pp.129–137, Mar. 1982.

[15] "Still images and sequences," Centre for Image Processing Research, Rensselaer Polytechnic Institute. [Online]. Available: http://www.cipr.rpi.edu/

[16] M. Arabzadeh, M.S. Helfroush, H. Danyali, and K. Kasiri, "Reversible watermarking based on generalized histogram shifting," in *Proc. IEEE ICIP*, 2011, pp.2797–2800.

[17] M. Fujiyoshi and H. Kiya, "Reversible data hiding for sparse histogram images," in *Proc. IEEK/IEICE/ECTI ITC-CSCC*, 2011, pp.510–513.