International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 1 (2019), pp. 42-49

Reversible Data Hiding for Medical Imagery Applications to Protect Privacy

Hae-Yeoun Lee

Dept. of Computer Software Engineering, Kumoh National Institute of Technology,
61 Daehak-ro, Gumi, Gyeongbuk, South Korea

ORCID: 0000-0002-6081-1492

Abstract

This paper presents a high quality and capacity data hiding technique for the medical imaging field which is recently known to have weak personal information protection. After analyzing the features of medical images, the data hiding technique is presented. To embed the private information, the technique calculates a histogram of the difference between the estimated value and the original value and then this histogram is expanded. The overflow and underflow possibility that can occur by information embedding is solved through an error-free scheme with location information and JBIG compression. This technique extracts the embedded information by inversing the embedding step and restores the original image without any modification. The performance of the presented technique is verified by analyzing the original image restoration rate. Also, since the information embedding capacity is determined according to the iteration count of the algorithm, embedding capacity and image quality are analyzed by increasing this count by 1 to 9 times. Therefore, this technique can be applied to medical imaging scanner or information processing system to protect privacy.

Keywords: Data hiding, high capacity, high quality, medical images, histogram expansion, error estimation.

I. INTRODUCTION

To protect personal information which is recently known to have weakness, this paper presents a high quality and capacity data hiding technique for the medical imaging field. It can be applied to medical imaging scanner or information processing system to protect privacy. Especially, the features of medical images are analyzed, the quantitative difference analysis in comparison with previous reversible data hiding for general images is explained, and the reversible data hiding technique which is modified for medical applications is presented.

Many data hiding methods such as robust data hiding, fragile data hiding, and semi-fragile data hiding, have been studied for general images. However, most of them are degraded the quality of data-hidden images and cannot restore the original images. Reversible data hiding can restore the original image and can be applicable for medical applications.

Previous reversible data hiding studies for general images are attempted to maximize embedding capacity while minimizing quality degradation [1-2]. Histogram-based studies, one of representative methods for reversible data hiding, can increase the embedding capacity by repeated embedding to the image in which an information is already embedded. Although these methods can be applied to medical images, the medical image is slightly different from general images using common CCD sensors. Therefore, it is possible to improve the performance by designing a specific method suitable for medical images by analyzing it.

The main requirements of reversible data hiding are high embedding capacity and 100% reversibility. Also, it requires a high image quality, i.e., perceptual transparency after information embedding. Recently, the computational complexity of data hiding should not be high due to the use of mobile devices with limited resource and computational capability [1].

This paper applies a high capacity and quality reversible data hiding technique which uses a difference histogram expansion and error-free scheme for medical images [2-3]. First, a histogram of difference values between the original value and the estimated value using an interpolation method between adjacent pixels is calculated, and then a private information is embedded by expanding this histogram. The possibility of the overflow and underflow problems that can occur by information embedding is solved through an error-free scheme with location information and JBIG compression [4]. The embedded information is extracted by inversing this embedding step and the original image is restored without any modification.

Unlike general images, medical images are used by doctors to check or identify illnesses of patients. Therefore, although the original image can be restored, the image distortion or quality degradation through data hiding before restoration may cause serious problems. Therefore, this paper verified the quality degradation with the original image restoration performance of the presented technique. Also, since the embedding capacity of the information is determined according to the iteration count of the technique, embedding capacity and image quality are analyzed by increasing this count from 1 to 9 times.

This paper is organized as follows. In Sec. II, related works are described. In Sec. III, the features of medical imagery are presented and the way to embed and extract medical information is presented. The performance is analyzed in Sec. IV and Sec. V concludes.

II. RELATED WORKS

In general, reversible watermarking can be classified into four categories. The first one is to compress the image to make a space for information embedding. Capacity depends on the compression algorithm and content applied. The second one is applying transformations such as DCT, DFT, and DWT to transform the domain and then modifying those coefficients. However, its computational complexity is high. The third one is to expand the difference between pixels to embed the information, but the location map to locate the modified pixels should be embedded together with the information and hence it degrades capacity. The last one is to histogram to embed message and this paper consider this category for data hiding of medical images.

Ni et al. [3] used the valley and peak locations of the histogram and its capacity was determined by pixel frequency belonging to the peak. After dividing the image into blocks, Lin et al. [5] inserted information by modifying the histogram for each differential image. Tsai et al. [6] used a differential image composed of the difference between a fixed pixel and the remaining pixels. After dividing the images into sub-sampled images and calculating their differential histograms, Kim et al. [7] inserted message by shifting each histogram. To solve overflow and underflow possibilities, Honsinger et al. [8] applied modulo-256 operation, but it causes salt and pepper noises reducing visual quality. Fridrich et al. [9] suggested a cyclic-modulo scheme to reduce these noise components. This is not a suitable solution because it reduces the width of variation.

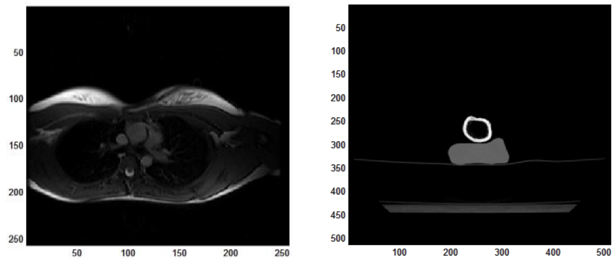
III. PRESENTED REVERSIBLE DATA HIDING

III.I Medical Image Feature Analysis

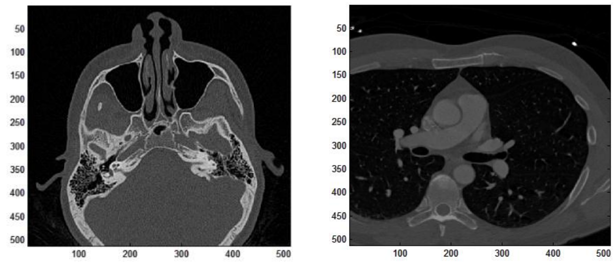
In the case of general images, they are 24 bits of 3 channels for the color image and 8 bits of 1 channel for the grayscale image. However, in the case of medical images, it corresponds to a grayscale image and generally has a 12-bit image quality. In other words, the intensity value of a pixel can have is 0 ~ 4091 and the range of intensity value change is very broad as compared with general images. In the case of general images, a method to measure the amount of visible lights that are detected during a specific moment by using CCD or CMOS sensors for shooting is used. However, in the medical images, there are various imaging modalities such as MRI, CT, and XRay. These modalities use a method to measure X-rays, not rays, or to measure the emitted energy. In addition, as shown in Fig. 1, most of the subjects are human bodies and they are characterized by a lot of similar parts to noise corresponding to the background and many parts corresponding to black which are not included in the object of interest.

III.II Reversible Data Hiding Technique

In the reversible data hiding method based on histogram expansion, a histogram is calculated, a maximum value is detected to embed an information, a histogram is modified based on this detected maximum value to make an empty space (or bin), and then the information is embedded by dispersing this maximum value depending the value of information. Therefore, in order to obtain a high embedding capacity, it is necessary to increase the maximum value of the histogram.



MRI-Cardiac image CT-Bone image



XRay-Brain image CT-Cardiac image

Fig. 1. Images depending on medical imaging modalities

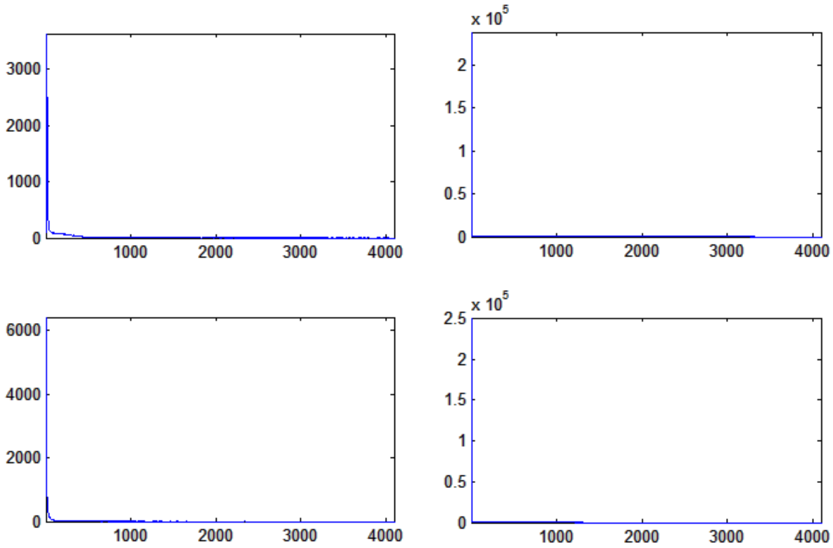
Initial histogram-based data hiding methods calculated this histogram using the pixel intensity value of the original image. In this paper, however, the method of calculating the histogram using the difference value between adjacent pixels is applied to increase the maximum value of the histogram.

Fig. 2 shows the intensity histogram in the spatial domain (top) and the difference histogram between adjacent pixels for various medical images (bottom). As shown in the figure, when the difference value between adjacent pixels is utilized, the probability of having the maximum value of the histogram to be high increases due to the locality of the pixel. Therefore, even if only one maximum bin in histogram is used, a high embedding capacity can be achieved. In particular, in the case of medical images, the profile of the histogram differs depending on target objects and imaging modalities.

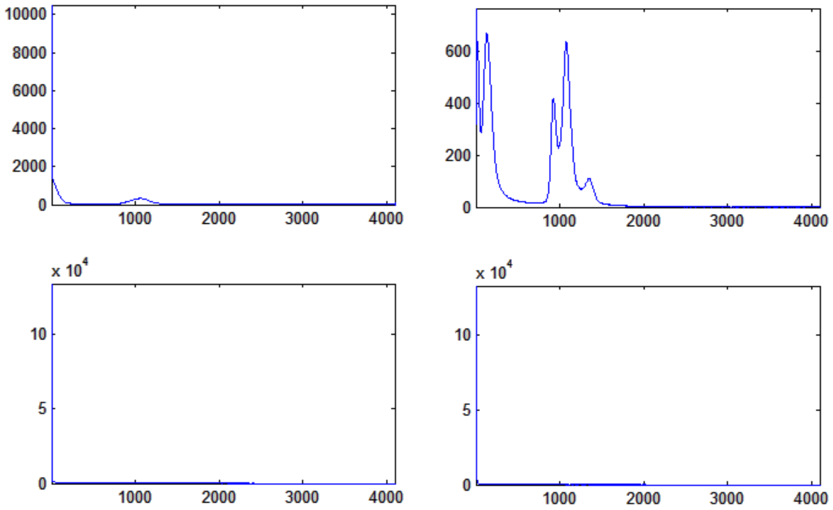
The method using this difference histogram can fix the position of the maximum histogram value due to the local characteristic of the pixel data which is small in the intensity change between adjacent pixels and assembled at zero value. Therefore, the overhead of the embedding location information can be eliminated.

III.III Private Information Embedding

The presented technique in this paper embeds the private information by changing the difference histogram between adjacent pixels. In general, the private information should be encrypted and hence the embedded information is assumed to be uniform in statistically. The process of private information embedding with error-free scheme to avoid the possibility of the overflow and underflow problem is schematically shown in Fig. 4.



MRI-Cardiac image CT-Bone image



XRay-Brain image CT-Cardiac image

Fig. 2. Intensity histogram and its difference histogram depending on medical imaging modalities and target objects

The presented technique can achieve a high embedding capacity by increasing the number of iterations R in the embedding steps according to the information capacity and image quality required. This information embedding step is composed of as following processing.

First, to avoid underflow and overflow possibility, an error-free scheme is performed for the original image I considering the number of times R to repeat the embedding process. As shown in (1), R is subtracted from the intensity value of the pixels whose the intensity value is larger than 65535-R in the spatial domain. Also, when the intensity value is smaller than R, R is added to the intensity value of the pixels. The modified pixel location is marked on the error-free location map.

$$I_e(i,j) = \begin{cases} I(i,j) - R & if\ I(i,j) >= 65535 - R \\ I(i,j) + R & if\ I(i,j) < R \\ I(i,j) & otherwise \end{cases} \tag{1}$$

where 0≤i<M, 0≤j<N and M and N represents the size of image.

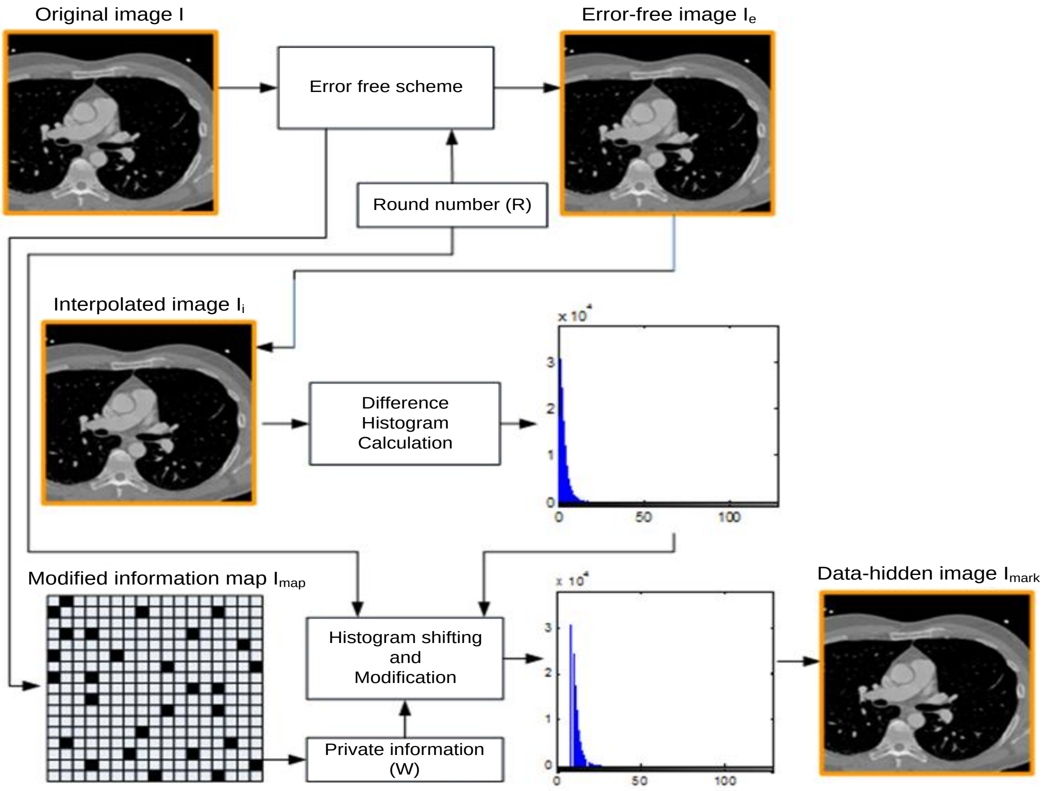


Fig. 3. Private information embedding steps of reversible data hiding

When the error-free scheme is performed, the error-free image I_e and the modified pixel information map I_{map} are generated. In order to construct an efficient data hiding system, this error information map is embedded into the image as part of the private information by performing compression.

To achieve maximum embedding capacity, the difference histogram between adjacent pixels is calculated. Using the error-free image I_e, the absolute difference image $D(i, j)$ is acquired by calculating difference with the interpolated image I_i. Then, the difference histogram H is generated from this absolute difference image $D(i, j)$.

The private information is embedded by shifting the difference histogram H. A position MAX_i corresponding to the maximum value is searched and values (bins) having a value larger than MAX_i is shifted by increasing its value as follows.

$$H_s = \begin{cases} H+1 & if\ H > MAX_i \\ H & otherwise \end{cases} \tag{2}$$

In the case of repetition count $R = 1$, there is a high probability that the maximum value exists at the zero position of the difference histogram. However, since the maximum value may exist at different positions depending on the image features, and hence the number of repetition times R should be sent additionally to the detection step.

After information embedding by shifting histogram, the pixel whose difference corresponds to the maximum value MAX_i of the histogram is searched to embed the information. When the embedding value is 0, the difference value is maintained. When the embedding value is 1, the difference value is increased by +1. This shifting and embedding can be modeled as follows.

$$D_s(i,j) = \begin{cases} D(i,j) & if\ D(i,j) < MAX_i \mid D(i,j) > MAX_i \\ D(i,j) & if\ D(i,j) = MAX_i \ \&\ W(n) = 0 \\ D(i,j)+1 & if\ D(i,j) = MAX_i \ \&\ W(n) = 1 \end{cases} \tag{3}$$

where $D(i, j)$ is the absolute difference image including the difference value, $0 \le i < M$, $0 \le j < N$ and M and N represents the image size. The embedded information W has a value of 0 or 1, and n is an index indicating the information position. This embedding process is repeated according to the specified number of repetitions R to embed the information.

After private information is embedded by histogram shifting and information embedding into the absolute difference image $D_s(i, j)$, the data-hidden image I_{mark} can be acquired by modifying the image pixel value depending on the difference value. The difference value is calculated by applying absolute operation, the directionality should be considered when expanding the difference value. This process is modeled as follows.

$$I_{mark} = \begin{cases} I_{cp}(i,j) - D_s(i,j) & if\ I_{co}(i,j) \ge I_{cp}(i,j) \\ I_{cp}(i,j) + D_s(i,j) & otherwise \end{cases} \tag{4}$$

where I_{co} corresponds to the reference point of the difference value calculation and I_{cp} is the target point of the difference value calculation.

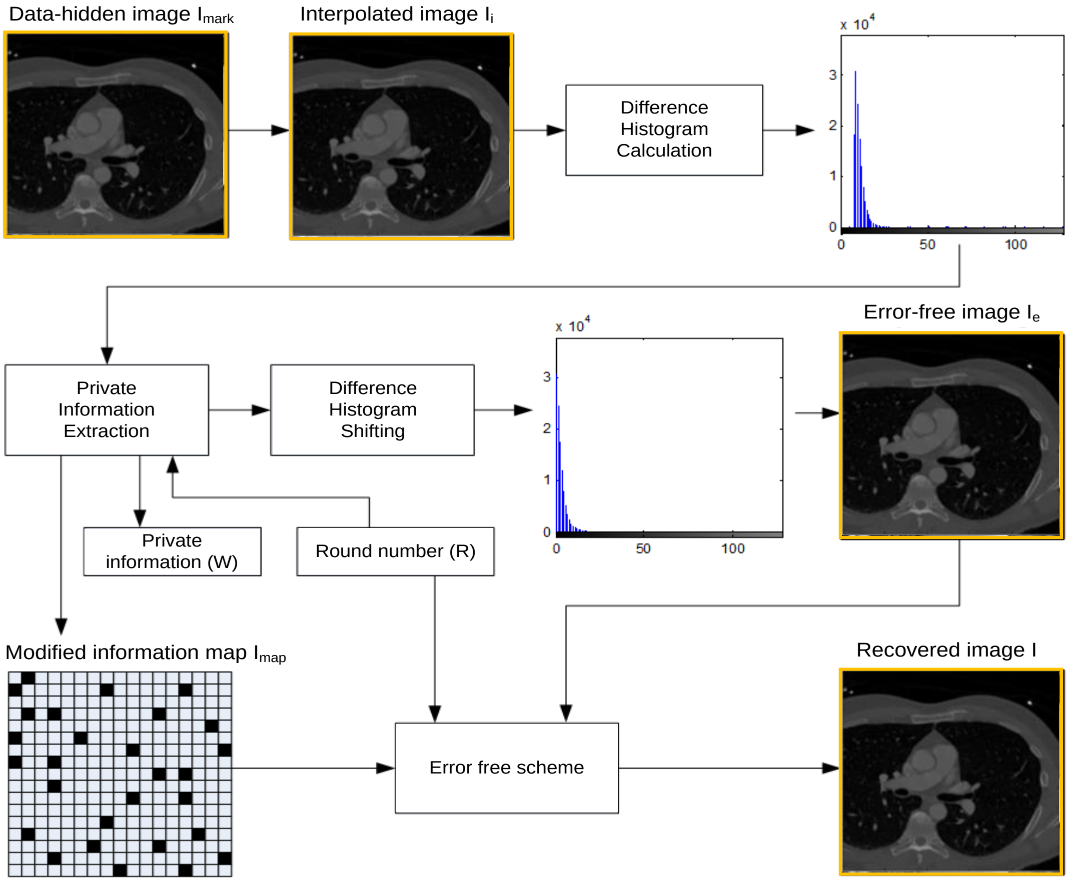


Fig. 4. Private information extraction and recovery steps of reversible data hiding

III.IV Private Information Extraction

The way to extract the hidden information is depicted in Fig. 4. To extract the hidden information, the number of repetitions R and the MAX_i information which is the positions of the histogram maximum values in each repetition are required. In case of these information, its size is a few bits and can be embedded into the image through the LSB bit plane compression method for transmission. Or, it can be transmitted additionally through another channel. This information extraction step is composed of as following processing.

First, the absolute difference image $D(i, j)$ is calculated using the data-hidden image I_{mark} and its interpolated image I_i similarly to the information embedding step. Then, the difference histogram H is generated from this absolute difference image $D(i, j)$.

The embedded private information is extracted using the MAX_i information about the position of the maximum histogram. This extraction process can be modeled as follows.

$$\mathrm{W(n)} = \begin{cases} 0 & if\ D(i,j) = MAX_i \\ 1 & if\ D(i,j) = MAX_i + 1 \end{cases} \qquad (5)$$

When the private information extraction is completed, the difference histogram H is restored to its original histogram H_o before the private information embedding through the process of histogram shifting as follows.

$$H_o = \begin{cases} H-1 & if\ H > MAX_i \\ H & otherwise \end{cases} \qquad (6)$$

This histogram shifting process has the same meaning as adjusting the difference value of the absolute difference image D as follows.

$$D_o(i,j) = \begin{cases} D(i,j) & if\ D(i,j) \le MAX_i \\ D(i,j)-1 & if\ D(i,j) > MAX_i \end{cases} \qquad (7)$$

where $D_o(i, j)$ corresponds to the absolute difference image from which the embedded information is removed.

When this absolute difference image D_o before the private information embedding is restored, it is possible to restore the error-free image I_e without including any information. This restoration process of the error-free image can be modeled as follows.

$$I_e(i,j) = \begin{cases} I_{wp}(i,j) - D_o(i,j) & if\ I_{wo}(i,j) \ge I_{wp}(i,j) \\ I_{wp}(i,j) + D_o(i,j) & otherwise \end{cases} \qquad (8)$$

where I_{wo} and I_{wp} is the reference point and the comparison point respectively to calculate the difference value in the data-hidden image I_{mark}.

The image from this restoration is the error-free image I_e rather than the original image I. In order to calculate the image I before error-free scheme, the number of repetitions R and the modified pixel location map I_{map} are required. Since the location map I_{map} was embedded and transmitted as a part of the private information and the repetition number R was transmitted as additional information, the original image can be restored through the following calculation.

$$I(i,j) = \begin{cases} I_e(i,j) + R & if\ I_{map}(i,j) = 1\ \&\ I_e(i,j) \geq 32 \\ I_e(i,j) - R & if\ I_{map}(i,j) = 1\ \&\ I_e(i,j) < 32 \\ I_e(i,j) & otherwise \end{cases} \quad (9)$$

The pixel which was marked as having a probability of overflow or underflow problems in the modified pixel location map can be perfectly restored by increasing the intensity value by the number of repetitions, decreasing the intensity value, and keeping the intensity of other pixels intact. As a result, the complete restoration of the original image I is possible.

IV. PERFORMANCE ANALYSIS

The presented data hiding technique was analyzed by using MRI, CT, and XRay images, which are widely used in the medical field as shown in Fig. 1. Also, images taken various subjects such heart, brain, and bone are considered for the experiment.

Since the embedding capacity is determined according to the number of repetition of the technique, the embedding capacity and the image quality are analyzed by increasing the number of repetition times R from 1 to 9 times. The embedded information is randomly generated so that 0 and 1 are evenly distributed based on the uniform distribution.

In order to calculate the embedding capacity, the effective embedding capacity was calculated by subtracting the size of the overhead information. Also, the image quality was measured using SNR and PSNR values. In particular, medical images are 12 bits or 16 bits, but display devices in practice can only output 8 bits (except for color conversion), so the analysis of PSNR quality through 8 bit conversion will be meaningful.

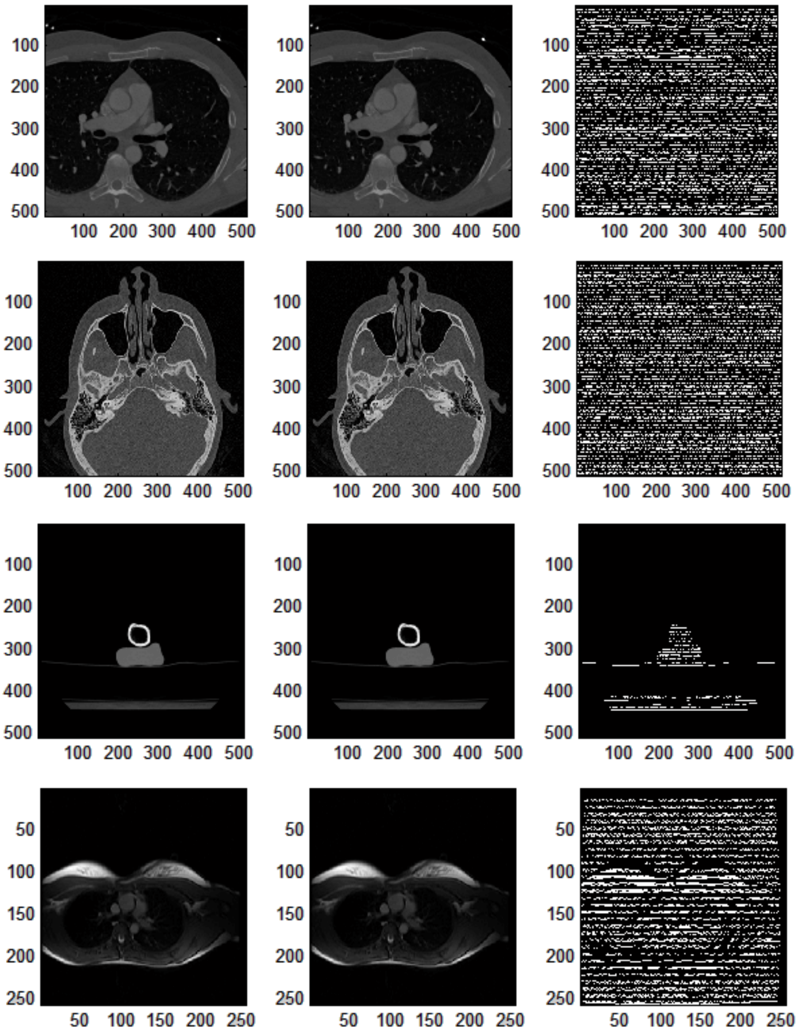
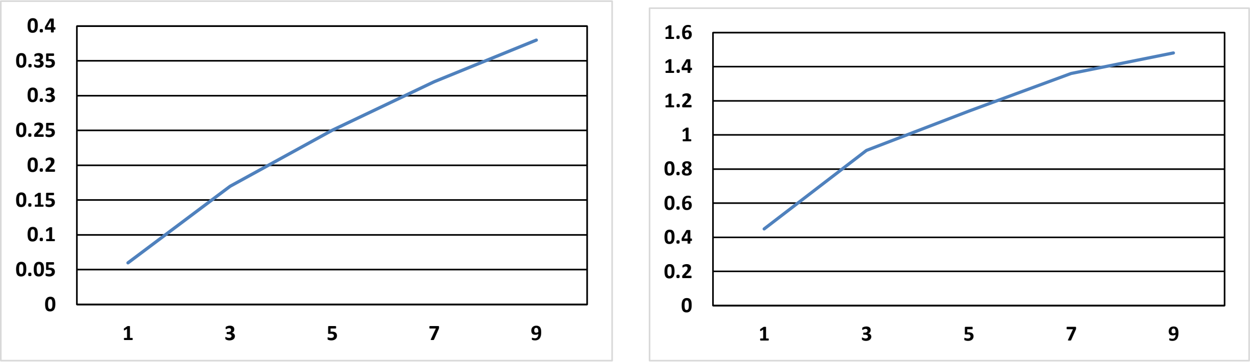
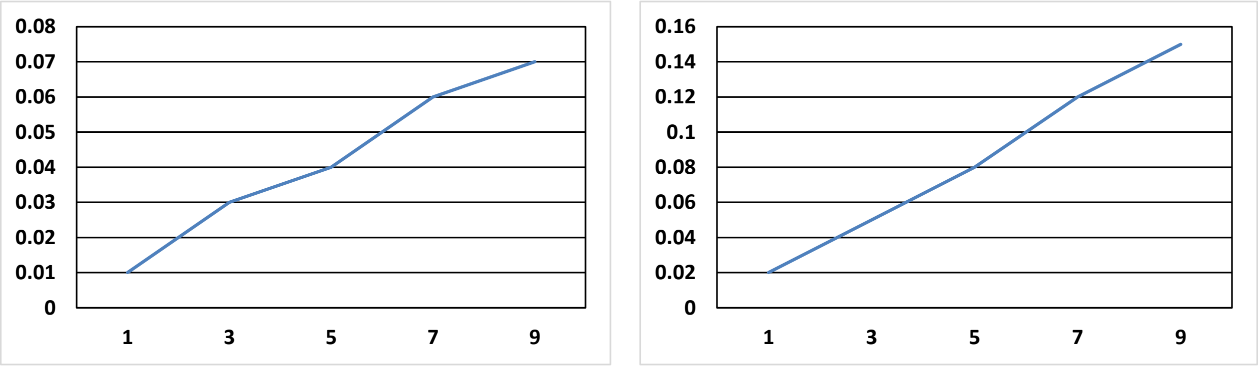


Fig. 5. Original image, data-hidden image, and their difference image after converting to 8 bits



MRI-Cardiac image

CT-Bone image



XRay-Brain image

CT-Cardiac image

Fig. 6. Embedding capacity of each image

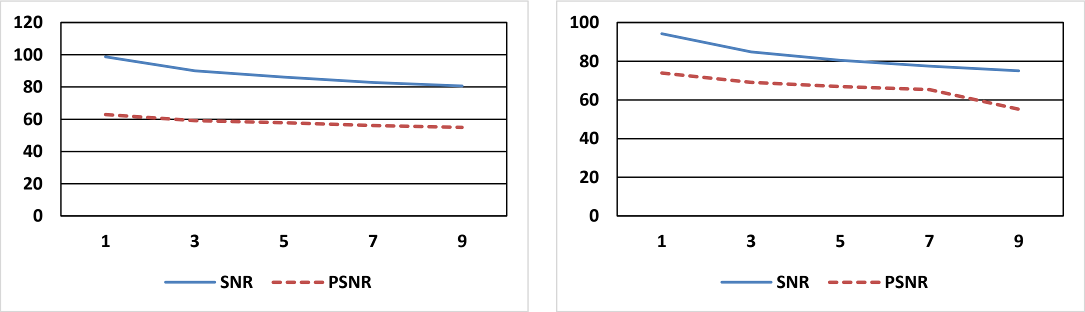
Fig. 5 shows the original image, the data-hidden image, and their difference value image after 8-bit conversion for various medical images. Qualitatively, it can be confirmed that the difference between the original image and the data-hidden image cannot be recognized. Especially, in the case of the CT bone image, there was no difference in the background part through private information embedding when 8 bit conversion is performed and the difference existed only in the bone part.

In Fig. 6 and Fig. 7, the embedding capacity and the image quality of each image according to the change of the number of repetition *R* are depicted. It can be confirmed that the embedding capacity increases as the number of repetitions increases. In particular, in the case of the CT bone image, it can be seen that the embedding capacity increases sharply due to the large number of background portions. Also, the SNR has a very high value and the PSNR after 8-bit conversion exceeds 54 dB, which confirms that the quality of the data hidden image is not different from that of the original image.

V. CONCLUSION

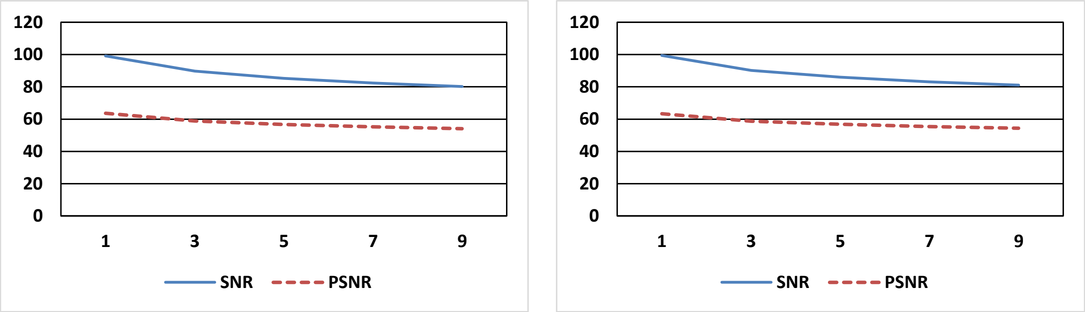
In this paper, a reversible data hiding technique based on histogram modification is presented to apply to medical images, which achieves high embedding capacity and high image quality, and low computational complexity. The features of medical images were analyzed, modified reversible data hiding technique was designed based on this features, and its performance was analyzed using medical images from various medical imaging modalities.

Since the presented histogram-based reversible data hiding is performed in the spatial domain, it can have a low computational complexity. Also, the private information embedding is performed by shifting the histogram, the error of +1/-1 occurs only depending on the number of embedding. As a result, it can achieve a high perceptual quality. The possibility of overflow and underflow problems that can occur in the histogram shifting was solved through error-free scheme.



MRI-Cardiac image

CT-Bone image



XRay-Brain image

CT-Cardiac image

Fig. 7. Perceptual quality of each image

Acknowledgements

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1 A1B03030432).

REFERENCES

[1] D. G. Yeo, H.-Y. Lee, and B. M. Kim, High Capacity Reversible Watermarking using Differential Histogram Shifting and Predicted Error Compensation, Journal of Electronic Imaging, 20(1), 2011, 013001-1-12.

[2] H.-Y Lee and K.-S Kim, High-Capacity Reversible Watermarking through Predicted Error Expansion and Error Estimation Compensation, KIPS Transactions on Software and Data Engineering, 17-B(3), 2010, 275-286.

[3] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, Reversible data hiding, IEEE Trans. on Circuits and Systems for Video Technology, 16(3), 2006, 354-362.

[4] M. Kuhn, JBIG-KIT lossless image compression library, (http://www.cl.cam. ac.uk/~mgk25/jbigkit/)

[5] C.-C. Lin, W.-L. Tai, and C.-C. Chang, Multilevel reversible data hiding based on histogram modification of difference images, Pattern Recognition, 41(12), 2008, 3582-3591.

[6] P. Tsai, Y.-C. Hu, and H.-L. Yeh, Reversible image hiding scheme using predictive coding and histogram shifting, Signal Processing, 89(6), 2009, 1129-1143.

[7] K.-S. Kim, M.-J. Lee, H.-Y. Lee, and H.-K. Lee, Reversible data hiding exploiting spatial correlation between sub-sampled images, Pattern Recognition, 42(11), 2009, 3083-3096.

[8] C. W. Honsinger, P. Jones, M. Rabbani, and J. C. Stoffel, Lossless recovery of an original image containing embedded data, US Patent application, Docket No: 77102/E-D, 1999.

[9] J. Fridrich, M. Goljan, and R. Du, Invertible authentication, in Proc. of the SPIE, Security and Watermarking of Multimedia Contents, San Jose, CA, 4314, 2001, 197-208.