Data Hiding Method with Principal Component Analysis and Image Coordinate Conversion

Kohei Arai

Faculty of Science and Engineering
Saga University, Saga City, Japan

Abstract—**Data hiding method with Principal Component Analysis (PCA) and image coordinate conversion as a preprocessing of wavelet Multi Resolution Analysis (MRA) is proposed. The method introduced in this paper, based on the characteristics of the original multispectral image, allows recovering the secret data. Through experiments, it is found that the proposed method is superior to the conventional data hiding method without any preprocessing. The method introduced in this paper allows only I who knows the characteristics of the original multispectral image to recover the secret data, i.e., when the information of the original image needs to be protected. Moreover, in the introduced method, the information of the secret data is protected by the existence of the eigenvector and the oblique coordinate transformation, that is, the secret data is restored if at least the information of the true original image is not known. The principal component transformation coefficient differs for each original image and is composed of the eigenvectors of the original image.**

Keywords—*Multi-dimensional wavelet transformation; multi resolution analysis (MRA); image data hiding; secrete image; Daubechies basis function*

I. INTRODUCTION

There are social problems such as illegal copying of digital contents such as DVDs and billing for music broadcasting / broadcasting. To solve this, it is necessary to keep content IDs and digital signatures highly confidential. Therefore, a global standard method such as copy prohibition and one-time copy permission is about to be developed. Digital contents of corporate electronic records, customer information, intellectual property, electronic medical records, etc. based on ISO 15489 (record preservation management guidelines) For copyright protection of digital contents (detection of digital signatures and tampering) that must be protected as personal information, and also by inserting a time tag in multimedia such as video, still image, music, etc. and using it to edit In order to efficiently perform charging, or to extract billing information classified by broadcast format, and only those who have obtained permission using a digital signature can share the content. Data Hiding techniques for such Kill manner are used [1].

Data hiding, which is also called information hiding, is a technique for hiding some information in the content. Here, the name of data hiding is used. Data hiding is a watermark (Digital Watermark) technique or steganography (Steganography) [2 - 4]. A technology that makes embedded information important and its existence unknown is called steganography. If the content itself in which confidential information is embedded is important, a digital watermark is used. The original content is called the original content, the data to be hidden such as a signature is called secret data, and the content in which the secret data is embedded is called distribution content. Watermark information embedding technology (watermark technology) has already been studied [5 - 8].

As a method of embedding the secret data in the content, a method of embedding the secret data in the real space of the content [9] and a method of embedding the secret data in the frequency space have been proposed [10],[11]. The latter has a higher ability to conceal the secret data information than the former because the secret data can be embedded in a specific frequency band that is relatively unaffected by the content quality. Therefore, for example, when the content is an image, it is necessary to manipulate the edge part of the original image to embed the secret data [9]. In the latter case, the frequency band of the original content in which the secret data should be embedded must be determined [7], [10].

A data hiding method has also been proposed when the original content is a color image [11], [12]. The secret data embedded in the original content is information such as copyright, and it has high resistance to image processing and removal attacks, and high confidentiality. The method of selecting the coefficient after the original content such as an image is transformed into the orthogonal frequency space and embedding the secret data in the frequency space is often used because it satisfies these requirements. Among them, a method has already been proposed to obtain distribution contents (images) by dividing the image into frequency components by wavelet multiresolution analysis (MRA) [13],[14], replacing any of the divided frequency component images with secret data, and reconstructing (combining).

The conventional method does not necessarily have sufficient confidentiality (it is possible to find it because any of the frequency component images contains confidential data), and the invisible (difficulty of visibility) of the confidential data is insufficient. When the original content is an image, data hiding using a color image has a higher ability to conceal secret data than other methods from the viewpoint of the amount of information in the original image. Therefore, a method of embedding secret data in a certain component (red, blue, or green component) of the original content (image) is generally used [4]. Since the embedding method is used, in this case, the information on the red and the blue components of the original image are not used. Because a certain color component is used in the embedding process.

Data hiding based on wavelet multi-resolution analysis is a method of investigating wavelet frequency components and embedding secret data in components that have a relatively small effect on image quality, but it is widely used. Since confidential data can be located by using this method, a problem remains in confidentiality. To overcome this problem, a method of applying principal component conversion to the original content (image) as a preprocessing of data hiding has also been proposed [15],[16],[17]. Since only I who owns the original content (image) can know the unique value of the original content (image), only I can restore it.

However, this method is not sufficiently confidential because it can estimate the approximate value of the eigenvalue by allowing a certain amount of error using the distribution contents (images) in which secret data is embedded. Pre-processing for cross coordinate transformation has also been proposed [18]. The cross angle of the cross coordinate can be set arbitrarily, and the principal component conversion can be performed based on this angle, which improves confidentiality [19]. A method has also been proposed that improves the visibility and confidentiality of confidential data in distribution images by converting the method [20].

Fundamentals of wavelet analysis and its application to data hiding are described in the books [21],[22],[23]. Method for data hiding based on Legall 5/2 (Cohen-Daubechies-Feauveau (CDF) 5/3) wavelet with data compression and random scanning of secret imagery data is proposed [24]. Improvement of secret image invisibility in circulation image with Dyadic wavelet-based data hiding with run-length coding is also proposed [25]. Meanwhile, noble method for data hiding using Steganography Discrete Wavelet Transformation (DWT) and Cryptography Triple Data Encryption Standard (DES) is proposed and well reported [26].

This paper outlines data hiding methods based on the wavelet multi-resolution analysis and evaluates the effect using images that are frequently used as standard images for data compression.

II. DATA HIDING BASED ON MULTI-RESOLUTION ANALYSIS

A. Wavelet Multi-Resolution Analysis

The biorthogonal wavelet decomposition (discrete wavelet transform based on biorthogonal basis function) applies $y = Cn$ x by applying the square matrix Cn to the original data (one-dimensional scalar data: $x = (x_1, x_2, ..., x_n)$). It can be defined as x, where Cn is a transformation matrix based on a biorthogonal basis function with $CnCn^t = I$. After conversion, y consists of low-frequency component L and high-frequency component H. That is, x is transformed as $y = (H1, L1)$, where the subscripts of H and L are the number of transformations, that is, the number of stages (level by applying Cn to this L1, it is transformed into H2 and L2, and by repeating this n stages, it is transformed into Hn and Ln. This is called decomposition.

The inverse transformation is $Cn^{-1} = Cn^t$ applying to y. By repeating this inverse transform n times, x is restored. This is called reconstruction. This wavelet transform / inverse transform (decomposition / reconstruction) is repeated Then, the decomposition into wavelet frequency components and the

reconstruction of the original data using the decomposed components are called multi-resolution analysis. When this is applied to two-dimensional data, for example, an image, $y =$ (HH1, HL1, LH1, LL1).

Here, HH1 means the high-frequency component in both the vertical and horizontal dimensions, and similarly LL1 means the low-frequency component in both the vertical and horizontal dimensions. This is called the two-dimensional wavelet transform. Each frequency component can be decomposed, and the original image data can be safely restored (reconstruction) by repeating the inverse transformation as in the case of one-dimensional data.

Fig. 1 shows the images obtained by applying the two-dimensional wavelet transform to the [Lenna] in the standard image database (SIDBA), which is often used to evaluate the data compression method, in one and two stages. At this time, Daubechies is used as the biorthogonal basis function. This original image is shown in Fig. 2.

When the DWT is applied to n time series data in one stage, it can be decomposed into n / 2 high frequency components and n / 2 low frequency components. By further subjecting the n / 2 low frequency components to a one-stage DWT, the n / 4 low frequency components and the n / 4 high frequency components can be decomposed.
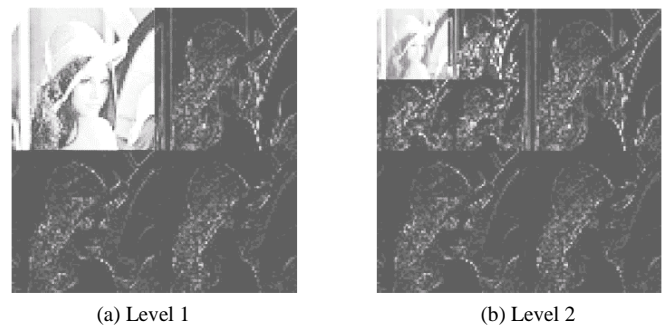

(a) Level 1   (b) Level 2

Fig. 1.   Examples of Images of which 1st and 2nd levels of DWT Applied to Lena of Original Image.
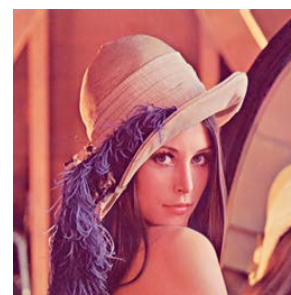

Fig. 2.   Original Image of Lenna.

B. Wavelet Multi-Resolution Analysis Based Data Hiding

The secret data embedded in the original content (image) is information such as copyright and signature (including the image), and data hiding requires resistance to image processing and removal attacks, and high confidentiality. The method of selecting the coefficient after transformation to the orthogonal frequency space and embedding the watermark in the frequency space is often used because it satisfies these

requirements. Among them, the wavelet multiresolution analysis introduced here (for each frequency component of the image This method is often used to obtain distribution data (images) by dividing them into images, replacing any of the divided frequency component images with secret data or secret data (images, signatures), and reconstructing (synthesizing).

C. Proposed Data Hiding Method

In order for improvement of confidentiality and visibility by eigenvalue expansion, the following data hiding method is proposed. Data hiding based on MRA is insufficient in confidentiality and visibility, and in order to overcome this problem, a method of performing principal component conversion as a preprocessing of data hiding based on MRA has been proposed. Only I who owns the original image can know the eigenvalue and eigenvector of the image, so only I can restore the original image. Therefore, I can claim the copyright of the original content. However, this method is not sufficiently confidential because the approximate value of the eigenvalue can be estimated by allowing a certain amount of error using the distribution image with embedded secret data.

In order to overcome this problem, there is also proposed a pre-processing to perform the oblique coordinate transformation after the principal component transformation, in which the oblique angle of the oblique coordinate can be set arbitrarily, and the principal component transformation can be performed based on this angle. The confidentiality of the content is highly protected under the condition that only the transmitting and receiving parties of the original content can know this angle information. Moreover, the Least Significant Bit (LSB) of the original content is encrypted by encrypting the angle information by a common key method. By inserting it in (the least significant bit in the quantization), the confidentiality can be set even higher. Here, I introduce data hiding based on multi-resolution analysis with principal component transformation and oblique coordinate transformation.

The process flow of this method is shown in Fig. 3. First, the energy of the original image is concentrated by principal component transformation, and the Cartesian coordinates of the transformed principal component image are transformed to oblique coordinates to further increase the energy concentration. Then, the MRA is applied to this, and the distribution image is obtained by reconstructing after embedding secret data in any of the levels and frequency components after decomposition. Since the principal component transformation parameters consisting of the eigenvalues and eigenvectors of the image are known, the original image and the secret data can be restored, but it is difficult for a third party who cannot know them to change the component to insert the secret data. This makes it possible to enhance the ability of data hiding based on multi-resolution analysis to protect information in secret data.

Cartesian coordinates and oblique coordinate representations in a two-dimensional plane are

$$W = X + Y \cos(\theta) \qquad (1)$$

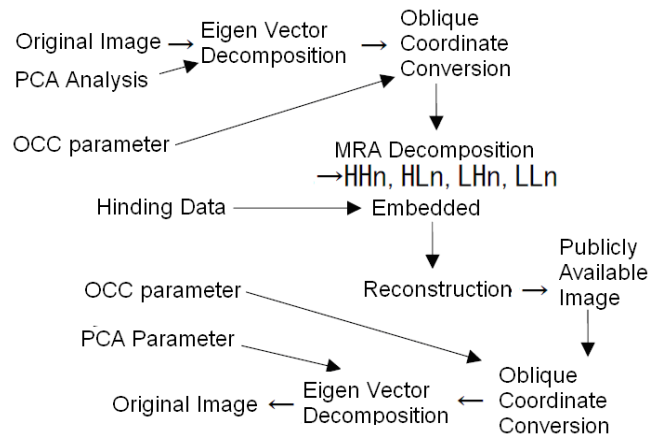$$Z = Y \sin(\theta) \qquad (2)$$

Fig. 3.   Process Flow of the Data Hiding based on MRA using Eigen Vector Decomposition (or Principal Component Analysis (PCA)) and Coordination Conversion from Cartesian to Oblique Coordination Systems.

where W, Z is each axis in the oblique coordinates, XY is each coordinate in the Cartesian coordinates, and θ is the angle of the coordinate axes in the oblique coordinate transformation. An example of this is shown in Fig. 4.

In the figure, the red and green two-dimensional pixel distributions (scatter diagram) of the original image are transformed into two-dimensional coordinates composed of the first principal component axis PC1 and the second principal component axis PC2 orthogonal to it. Also, the Cartesian coordinates consisting of PC1 and PC2 are converted to the diagonal coordinates of the diagonal angle θ consisting of PC1 'and PC2'.

At this time, if the diagonal angle is changed without changing the quantization step, pixel definition is performed. The extreme example is when the domain is below the quantization step. In this case, only the quantization noise is transmitted, and no information is transmitted. The domain can be expanded (e.g. the pixel value is doubled when changing the oblique angle in the range of 90 degrees ± 45 degrees) and reduced when restoring the original image and the secret data. A reversible process, and this enlargement / reduction rate is known by only for content owners. Therefore, the only content owners can fully restore the original image.
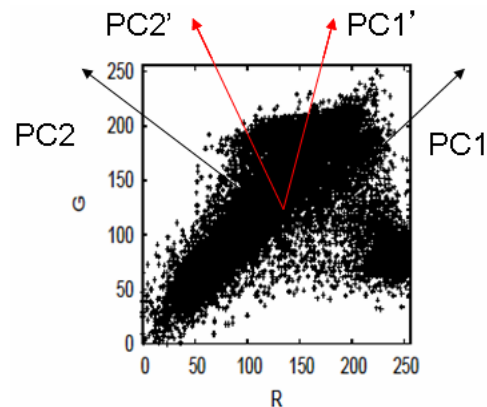

Fig. 4.   PCA and Oblique-Coordinate Conversion (OCC) as a Preprocessing of the MRA based Data Hiding.

Next, the method of decrypting the secret data is explained. The first principal component image is used for the distribution image by using the coefficient when the principal component conversion is applied to the multidimensional original image before the secret data is hidden. Is implemented and wavelet decomposition is performed on the first principal component image. Decoding the secret data by the proposed method transforms the principal component transform into the multidimensional original image before hiding the secret data information. Decoding is possible only when the coefficients used are known, that is, the coefficients of principal component transformation differ depending on the multidimensional original image before hiding the secret data.

III. EXPERIMENTS

An example of the experiment is shown as follows. I also used the Mandrill (Fig. 5), which was also selected from the standard image database for data compression evaluation (SIDBA), as the original image, and the time series data shown in Fig. 6 as the secret data. (Graph) was used.

The red, green, and blue primary color images of this color primary image are shown in Fig. 7. Here, the blue component is set to 0, and for the sake of convenience, it is used as two-dimensional multispectral image data. Fig. 8 shows the scatter diagram of the mean vector and transform coefficient matrix for performing the principal component transformation from the red and green two-dimensional scatter of the original color image, that is, the eigenvalue and the eigenvector, respectively.

λ =(133.772, 129.297)

V=|0.835   0.550|

|− 0.550 0.835|

Only I who owns the original image knows these accurate eigenvalues and eigenvectors (principal component conversion parameters), and even if the eigenvalue expansion is performed based on the circulation image, the restored image when the correct eigenvalues / eigenvectors are used. However, if the confidential data is image data with high redundancy, it is possible to recover the confidential data within the allowable error range and it is not enough. In order to enhance the confidentiality, I decided to add the oblique coordinate transformation as a pre-processing.

Convert the Cartesian coordinate axes after principal component conversion to any diagonal coordinate axes using the above Eqs. (1) and (2), where the diagonal angle of the diagonal coordinate axes is the diagonal coordinate transformation parameter θ. θ = 90 degrees is the orthogonal coordinate axis itself after the principal component conversion, and for example, if oblique coordinate conversion is performed with θ = 110 degrees and 70 degrees, Fig. 8 becomes Fig. 9. This diagonal angle can be arbitrarily set up to the range of the quantization step as described above, but the scaling ratio of the pixel value domain is changed depending on this angle. Therefore, it is necessary to adjust the quantized bit accordingly, and it is necessary to increase the processing resource, so I set it here to about ± 20 degrees. Restoration is extremely difficult because there is no information such as the diagonal angle and eigenvalue vector.
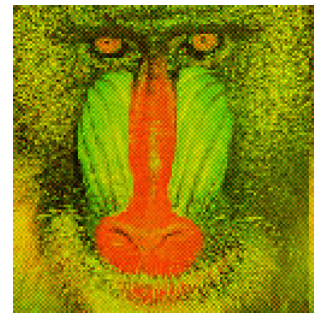

Fig. 5.   Original Image of Mandrill of SIDBA Database for Data Compression Algorithm Evaluation.
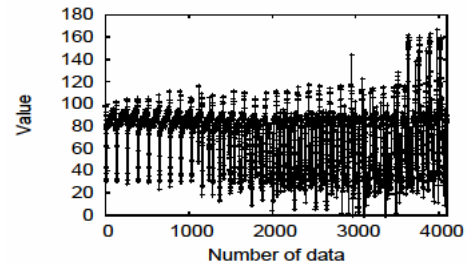

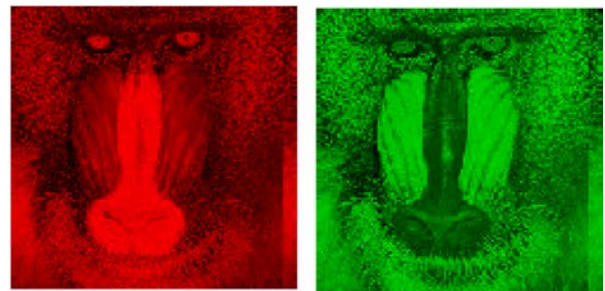Fig. 6.   Hiding Data of Image (Time Series of Data Graph).

Fig. 7.   Red and Green Colored Images of the Original Colored Image of Mandrill (Fig. 5).
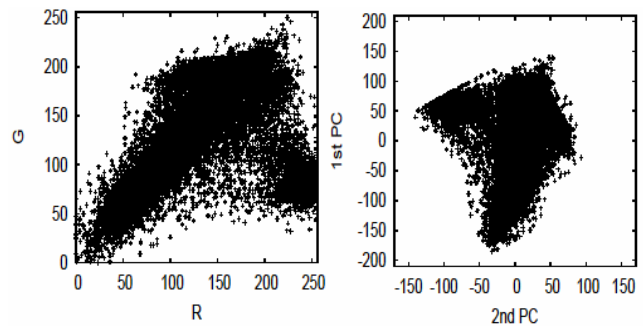

Fig. 8.   Scatter Diagrams of the Original Red and Green Coordinate System and that of the First and Second Components Coordinate System.

It is shown that the protection performance of the secret data is improved by the parameter θ. Therefore, I try to estimate the secret data from the first principal component image in Fig. 10 by using the wavelet transformation. The information of the wavelet basis and the component (for example, the HH1 component) in which the secret data is embedded is known by some method, i.e., the information such as the eigenvectors held by the parties and the parameter θ are known. The third party is unknown.

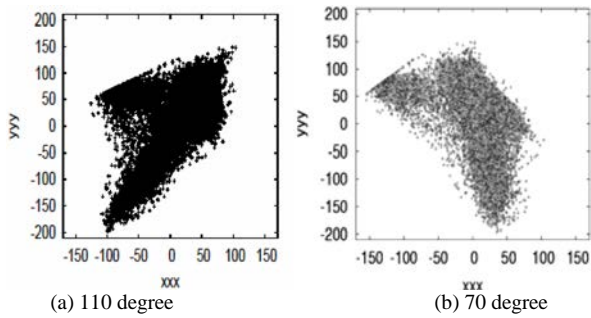(a) 110 degree   (b) 70 degree

Fig. 9.   Oblique Coordinate Converted Scatter Diagrams with 110 and 70 Degrees.

Fig. 10.   Widely Available Image of which Hiding Image (Time Series of Data Graph) is hidden in the Original Image (Mandrill).

Fig. 11 shows the RMS deviations when secret data is embedded by changing the oblique coordinate transformation parameter θ and a third-party attempts to estimate secret data from the distribution data for each distribution data.

The RMS deviation when a third-party attempts to estimate the secret data from the distribution data depends on θ, that is, θ increases the confidentiality of the secret data. Therefore, it can be seen that the confidential data can be protected by protecting the information of the original image.

Next, I examine the degree of restoration of the secret data. I restored the secret data from the distribution image and evaluated the degree of restoration (mean squared error (RMS error) between the secret data and the restored secret data. The results are shown in Fig. 12. From this figure, it can be seen that the mean squared error between the original secret data and the secret data reconstructed from the distribution image increases. In other words, increasing the diagonal angle improves the confidentiality.

It is important for hiding. The degree of restoration of secret data was evaluated when noise due to image processing, removal attacks, etc. was included in the distribution image. The noise included was 5 steps with a mean of 0 and a standard deviation sigma of 5-20. After superimposing the normal random number changed in step 1, the secret data was reconstructed and the mean square error from the original secret data was evaluated. The results are shown in Fig. 13.

The mean square error increases steadily by the amount of superimposed noise, and the mean square error increases by about 10% by changing the oblique coordinate angle from 90 degrees (orthogonal coordinates) to 110 degrees. Therefore, there is no resistance to noise, and it is found that the restored secret data deteriorates by the amount of superimposed noise,

and the effect of oblique coordinate transformation is about 10%. Although it is possible, the quantization error due to requantization after coordinate transformation increases, so it was judged that the limit is about 70 to 110 degrees.
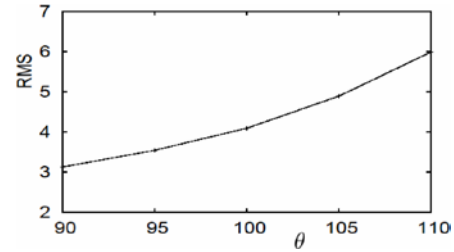

Fig. 11.   Root Mean Square difference between Original and Distributing Image of the Proposed Watermarking with 90 to 110 Degrees of OCC.
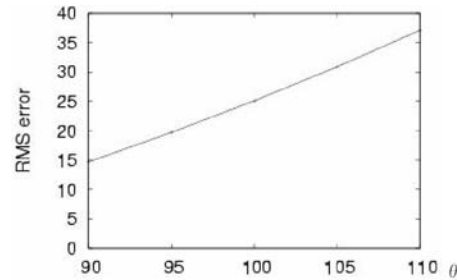

Fig. 12.   RMS Error between Original and Restored Hiding Data Derived from the Distributing Image.
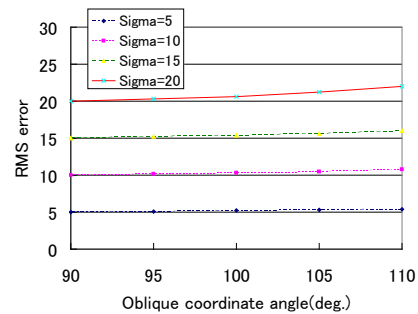

Fig. 13.   Hiding Performance for which Normal Distributed Noise with Zero mean and Standard Deviation Ranges from 5 to 20 (5 Step) is Added to the Distributing Image.

IV.   CONCLUSION

Data hiding method with Principal Component Analysis (PCA) and image coordinate conversion as a preprocessing of wavelet Multi Resolution Analysis (MRA) is proposed. The method introduced in this paper allows only I who knows the characteristics of the original multispectral image to recover the secret data, i.e., when the information of the original image needs to be protected. In this paper, the Daubechies basis function is adopted as the wavelet, but the secret data can be restored by using the biorthogonal wavelet, and the secret data can be protected by hiding what is adopted as the biorthogonal wavelet.

Through experiments, it is found that the proposed method is superior to the conventional data hiding method without any preprocessing. The method introduced in this paper allows only the owner of the original image who knows the characteristics

of the original multispectral image to recover the secret data, i.e. when the information of the original image needs to be protected. Moreover, in the introduced method, the information of the secret data is protected by the existence of the eigenvector and the oblique coordinate transformation, that is, the secret data is restored if at least the information of the true original image is not known. The principal component transformation coefficient differs for each original image and is composed of the eigenvectors of the original image.

I have introduced a method that improves the confidentiality by applying principal component transformation and oblique coordinate transformation as preprocessing for data hiding based on wavelet multiresolution analysis. I investigated the confidentiality when a third-party attempts to extract secret data from only the data for distribution.

The proposed data hiding method can be applicable for all the images in the world. There is no limitation in terms of applicability at all.

V. FUTURE RESEARCH WORKS

In the future, I will compare the proposed method with conventional data hiding methods such as steganography method. Influences on the restoration process due to tampering on the secret image hidden imagery data has to be investigated.

REFERENCES

[1] Tsukasa Ono, Digital Watermarking and Content Protection, Ohmsha, (2001).

[2] K. Matsui, Basics of Digital Watermarking, Morikita Publishing, (2000).

[3] Atsuro Kawai, Watermark-Authentic proof / paper scraper (watermark) The Institute of Image Electronics Engineers, 31, 2, 253-260, (2002).

[4] Eiji Kawaguchi, Hideki Noda, Michi Niimi, Digital steganography technology, IEICE, 31, 3, 414-420, (2002).

[5] Tanaka, K., Y. Nakamura and K. Matsui: "Embedding Secret Information into a Dithered Multilevel Image" Proceedings of the 1990 IEEE Military Communications Conference, 216-220, (1990).

[6] Tirkel, A., et al., "Electronic Water Mark", Proceedings DICTA 1993, 666-672, (1993).

[7] Cox, I.J., J. Killian, T. Leighton, T. Shamoon: "Secure spread spectrum watermarking for multimedia", NEC Technical Report 95-10, NEC Research Institute (1995).

[8] Bender, W. D. Gruhl, and N. Morimoto : "Techniques for Data Hiding" Proceedings of the SPIE 2420, Storage and Retrieval for Image and video Database III, 164-173, (1995).

[9] Akihiro Okamoto and Akio Miyazaki, Digital Watermarking Method Using Morphological Signal Processing, IEICE Transactions A, J84-A, 8, 1037-1044, (2001).

[10] Hisashi Inoue, Akio Miyazaki, Mikio Shimazu, Takashi Katsura, Digital Watermarking Method for Image Signals Using Wavelet Transform, Journal of Image Information Media, 52, 12, 1832-1839, (1998).

[11] Ouellette, R., H. Noda, M. Niimi, and E. Kawaguchi, Topological Ordered Color Table for BPCS Steganography Using Indexed Color Images, IPSJ Journal, 42, 1, 110-113, (2001).

[12] Seto, Hidetoshi, Morita, Yuji, Kang, and Aoki, Y., Color image watermarking method using Fresnel transform in image compression, IEICE Tech. Tech., 43-48 (2000).

[13] Kohei Arai, Basic Theory of Wavelet Analysis, Morikita Publishing, (2000).

[14] Kohei Arai,, Self-learning wavelet analysis, Modern Science Company, (2006).

[15] Kohei Arai, Kaname Seto, Data Hiding Based on Wavelet Multiresolution Analysis, Journal of Visual Information Society, 22, Suppl.1, 229-232, (2002).

[16] Kohei Arai, Kaname Seto, Data hiding based on multi-resolution analysis using bias of information by principal component transformation, Journal of Visual Information Society, 23, 8, 72-79, (2003).

[17] Kohei Arai and Kaname Seto, Information Hiding Method Based on Principal Component Analysis, Proc. Of the 37th conference of the Remote Sensing Society of Japan, (2004).

[18] Kohei Arai, Kaname Seto, Information hiding method based on coordinate transformation, Journal of Visual Information Society, 25, Suppl.1, 55-58, (2005).

[19] Kohei Arai and Kaname Seto, Data hiding based on biorthogonal wavelet multiresolution analysis considering scanning method, Proc. Of the 35th conference of RSSJ, (2003).

[20] Kohei Arai and Kaname Seto, Difficulty of key image recognition based on biorthogonal wavelet multi-resolution analysis considering scanning method, Data hiding with improved confidentiality, Journal of Japan Society for Visual Information, 29, Suppl. (2009).

[21] Kohei Arai, Basic Theory of Wavelet Analysis, Morikita Publishing (November 2000).

[22] Kohei Arai, Leland Jameson, How to use earth observation satellite data by wavelet analysis, Morikita Publishing (July 2001).

[23] Kohei Arai, Self-study wavelet analysis, published by Modern Science Co., Ltd. (June 2006).

[24] Kohei Arai, Method for data hiding based on Legall 5/2 (Cohen-Daubechies-Feauveau: CDF 5/3) wavelet with data compression and random scanning of secret imagery data, International Journal of Wavelets Multi Solution and Information Processing, 11, 4, 1-18, B60006 World Scientific Publishing Company, DOI: I01142/SO219691313600060, 1360006-1, 2013.

[25] Kohei Arai and Yuji Yamada, Improvement of secret image invisibility in circulation image with Dyadic wavelet based data hiding with run-length coding, International Journal of Advanced Computer Science and Applications, 2, 7, 33-40, 2011.

[26] Cahya Rahmed Kohei Arai, Arief Prasetyo, Noriza Arigki, Noble Method for Data Hiding using Steganography Discrete Wavelet Transformation and Cryptography Triple Data Encryption Standard: DES, IJACSA, 9, 11, 261-266, 2018.

AUTHOR'S PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Science Commission "A" of ICSU/COSPAR since 2008 then he is now award committee member of ICSU/COSPAR. He wrote 55 books and published 620 journal papers as well as 450 conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html