Secure Copier Which Allows Reuse Copied Documents with Sorting Capability in Accordance with Document Types

Kohei Arai [1]
Graduate School of Science and Engineering
Saga University
Saga City, Japan

***Abstract*—Secure copy machine which allows reuse copied documents with sorting capability in accordance with the document types. Through experiments with a variety of document types, it is found that copied documents can be shared and stored in database in accordance with automatically classified document types securely. The copied documents are protected by data hiding based on wavelet Multi Resolution Analysis: MRA.**

Keywords—data hiding; MPA; wavelet; secure copy

I. INTRODUCTION

The conventional copy machines can make a copy. Also the conventional secure copiers erase copied documents immediately after the documents are copied. The secure copy machine proposed here has the following functionalities,

1) Make a copy
2) Save the copied documents in database of classified document types after applying data hiding
3) Copied documents can be reused by among the authenticated users

Because the proposed copy machine is secure enough with data hiding, copied documents are protected from outsiders without authentication even if outsiders break network security and database security. Therefore, there is no need to erase the copied documents. Thus the proposed copy machine allows reuse the copied documents. Therefore, the copied documents have to classify by the document types for convenience of reuse. Key methods of the proposed secure copy machine are as follows,

1) Data hiding
2) Classification of copied documents

Data hiding method based on wavelet Multi Resolution Analysis: MRA has been proposed already [1]-[3]. Also the method for keyword extraction from the original documents based on Analytic Hierarchical Process: AHP [1] has been proposed already [4]. Invisibility of hidden documents has also been improved by means of random scanning [5].

[1] http://ja.wikipedia.org/wiki/%E9%9A%8E%E5%B1%A4%E5%88%86%E6%9E%90%E6%B3%95

The following section describes the proposed secure copy machine together with the key methods of data hiding and classification methods. Then experiments on data hiding and classification performance will be followed. Finally, conclusion with some discussions is described.

II. PROPOSED METHOD

The proposed secure copy machine is not only for making copy of original documents but also for making the copied documents available to use again for authenticated users. In this section, system configuration is, at first, described followed by process flow. Then the key components of the system, steganography and watermarking [2] for the copied documents, keyword extraction from the copied documents, classification of the copied documents are described.

A. System Configuration and Process Flow

System configuration of the proposed document server which allows reuse previously acquired documents which are protected with steganography and watermarking and watermarking is shown in Figure 1.

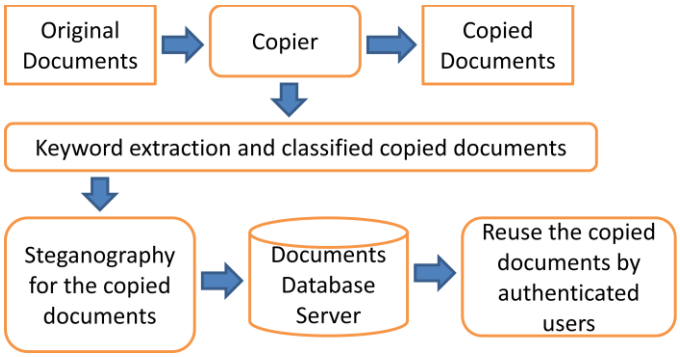

Fig. 1. System configuration of the proposed document server which allows reuse previously acquired documents which are protected with steganography and watermarking and watermarking

Original documents can be copied and use the copied documents for general users. Once the original documents are copied, keyword of the documents is extracted from the copied documents automatically and then classified the copied

[2] http://www.jjtc.com/Steganography/

documents based on the extracted keyword. After that, steganography and watermarking is applied to the copied documents and stored in the documents database server. Therefore, the copied documents can be reused by authenticated users who know the way to decryption of the steganography and watermarking of reusable copied documents.

B. Steganography and Watermarking

The copied documents are to be stored in the appropriate subject holder in the documents database server through steganography and watermarking. Even if the general users who do not authenticated at all access to the documents database server and acquire the copied documents, such users could not decryption the original documents at all because such users do not know the way for decryption of the documents which are protected by steganography and watermarking.

Original image can be decomposed with horizontally and vertically low as well as high wavelet frequency components based on wavelet Multi Resolution Analysis: MRA which is expressed in equation (1).

$$F=C_n\eta \quad (1)$$

Where F, C_n, and η is wavelet frequency component, wavelet transformation matrix, and input data in time and/or space domain. Because $C_n^t=C_n^{-1}$, η can be easily reconstructed with wavelet frequency component, F. Equation (1) is one dimensional wavelet transformation and is easily expanded to multi dimensional wavelet transformation.

$$F=(C_n\,(C_m\,(C_l\eta\,)^t)^t)^t... \quad (2)$$

In the case of wavelet transformation of images, two dimensional wavelet transformations is defined as follows. Horizontally low wavelet frequency component and vertically low frequency component is called LL_1 component at the first stage. Horizontally low wavelet frequency component and vertically high frequency component is called LH_1 component at the first stage. Horizontally high wavelet frequency component and vertically low frequency component is called HL_1 component at the first stage. Horizontally high wavelet frequency component and vertically high frequency component is called HH_1 component at the first stage. Then LL_1 component can be decomposed with LL_2, LH_2, HL_2, and HH_2 components at the second stage. Also LL_2 component is decomposed with LL_3, LH_3, HL_3, and HH_3 components at the third stage as shown in Figure 2.

Thus Laplasian pyramid[3] which is shown in Figure 3 is created. If these four decomposed components, LL_n, LH_n, HL_n, and HH_n are given, then LL_{n-1} is reconstructed. Therefore, the original image can be reconstructed with all the wavelet frequency components perfectly.

The copied documents can be replaced into the designated portion of wavelet frequency component. Furthermore, the Least Significant Bit: LSB is also replaced to the encrypted location of wavelet frequency component in which the copied documents are replaced. Moreover, the encrypted location is randomly scanned with Mersenne Twister[4] of random number generator. Therefore, only the authenticated users who know the parameters of the random number generator can decode the location of the wavelet frequency component.

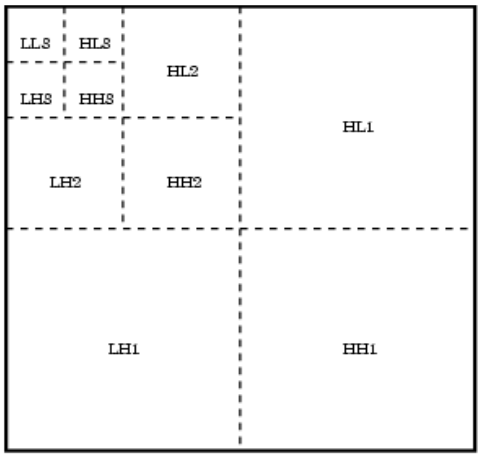

Fig. 2. Two dimensional wavelet transformation

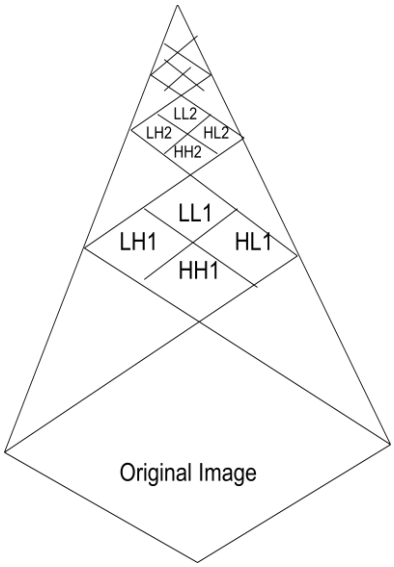

Fig. 3. Laplasian pyramid of wavelet Multi Resolution Analysis: MRA

C. Keyword Extraction

Documents can be classified into three categories, (1) figures which include photos, diagrams, drawings, illustrates, etc., (2) documents, and (3) tables. One of the examples of the figures is shown in Figure 4. These categories of documents have keywords in the documents. For instance, the example has its keywords which are located at "A" in the figure.

Based on the knowledge on the location, font size, frequency for keywords, the keywords can be extracted. For example, keywords are located at the top left, center, and right corners as well as the bottom left, center, and right corners. The

[3] http://en.wikipedia.org/wiki/Laplacian_pyramid

[4] http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/mt.html

font size of the keywords is relatively large in comparison to the other. The keywords are used to be appeared frequently. These features of keywords are common to the all kinds of documents types.

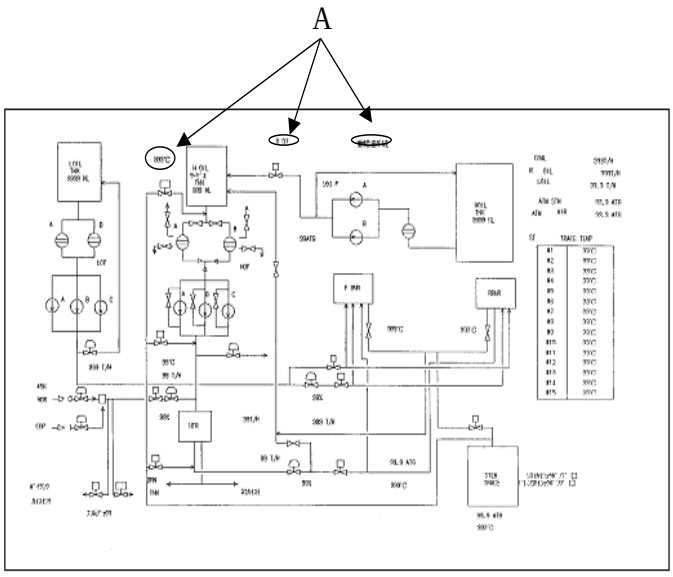

Fig. 4. Example of the documents of figures category

D. Classification of the Copied Docuemnts

The extracted keywords are categorized through Morphological Analysis[5] using "ChaSen"[6] through the URL : http://chasen.aist-nara.ac.jp.

Then the keywords are decomposed with Morpheme[7]. After that, lexical conceptual structure is checked for classify the copied documents. Thus the copied documents are categorized into the appropriate subjects and stored in the documents database server for reuse of the copied documents.

III. EXPERIMENTS

A. Steganography and watermarking

Using the forth order of Daubechies wavelet[8] base function of MRA, copied document (image) (see Appendix) is hidden (steganography and watermarking). One of the example images is shown in Figure 5. "Lena" (one of the standard image for data compression performance evaluation, SIDBA[9]) of Figure 5 (a) is decomposed with Figure 5 (b) and (c).

Figure 6 shows the example of which the copied document of "CRAMPS" is hidden in LH_1 wavelet frequency component. Also, Figure 7 shows the example of reconstructed image which contains the copied document of "CRAMPS".

(a)Lena image

(b)First stage of the decomposed image

(c)Second stage of the decomposed image

Fig. 5. Example of decomposed images

[5] http://en.wikipedia.org/wiki/Morphological_analysis
[6] http://ja.wikipedia.org/wiki/ChaSen
[7] http://en.wikipedia.org/wiki/Morpheme
[8] http://en.wikipedia.org/wiki/Daubechies_wavelet
[9] http://imagingsolution.blog107.fc2.com/blog-entry-180.html

Fig. 6. Example of which the copied document of "CRAMPS" is hidden in LH1 wavelet frequency component

The users who are not authenticated can get only the image of Figure 7. The copied document, "CRAMPS" cannot be gotten by such users. On the other hand, the authenticated users may get the copied document, "CRAMPS" because they know the parameters of random number generator, the location of wavelet frequency component, how to decryption of the code.

Fig. 7. Example of reconstructed image which contains the copied document of "CRAMPS".

IV. Conclusion

Secure copy machine which allows reuse copied documents with sorting capability in accordance with the document types. Through experiments with a variety of document types, it is found that copied documents can be shared and stored in database in accordance with automatically classified document types securely. The copied documents are protected by data hiding based on wavelet Multi Resolution Analysis: MRA.

Through the experiments with SIDBA image database, it is found that the proposed secure copier ensure that original image can be protected from unauthenticated users and can be accessible from authenticated user who knows initial parameter of the random number generator.

Appendix

Daubechies wavelet base function utilized wavelet transformation (or decomposition) is defined as follows,

$$C_n \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} \tag{A1}$$

Where η denotes input data while C_n denotes wavelet transformation matrix which is represented as follows,

$$C_8^{[2]} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \end{bmatrix} = \begin{bmatrix} p_0 & p_1 & & & & & & \\ q_0 & q_1 & & & & & & \\ & & p_0 & p_1 & & & & \\ & & q_0 & q_1 & & & & \\ & & & & p_0 & p_1 & & \\ & & & & q_0 & q_1 & & \\ & & & & & & p_0 & p_1 \\ & & & & & & q_0 & q_1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \end{bmatrix}$$

$$= \begin{bmatrix} p_0 x_1 + p_1 x_2 \\ q_0 \eta_1 + q_1 \eta_2 \\ p_0 \eta_3 + p_1 \eta_4 \\ q_0 \eta_3 + q_1 \eta_4 \\ p_0 \eta_5 + p_1 \eta_6 \\ q_0 \eta_5 + q_1 \eta_6 \\ p_0 \eta_7 + p_1 \eta_8 \\ q_0 \eta_7 + q_1 \eta_8 \end{bmatrix} \tag{A2}$$

Where n denotes input data size (or order) while p and q are determined with the following equation (A3),

$$\begin{aligned}\left(C_n^{[2]}\right)^T C_n^{[2]} &= I_n \\ p_0 + p_1 &= \sqrt{2} \\ q_0 &= p_1 \\ q_1 &= -p_0 \\ 0^0 q_0 + 1^0 q_1 &= 0\end{aligned} \tag{A3}$$

Equation (A2) is for the support length of two of wavelet transformation matrix while the support length of four of wavelet transformation matrix is expressed as follows,

$$C_8^{[4]}\begin{bmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\\ \eta_5\\ \eta_6\\ \eta_7\\ \eta_8\end{bmatrix} = \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & & & & \\ q_0 & q_1 & q_2 & q_3 & & & & \\ & & p_0 & p_1 & p_2 & p_3 & & \\ & & q_0 & q_1 & q_2 & q_3 & & \\ & & & & p_0 & p_1 & p_2 & p_3 \\ & & & & q_0 & q_1 & q_2 & q_3 \\ p_2 & p_3 & & & & & p_0 & p_1 \\ q_2 & q_3 & & & & & q_0 & q_1 \end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\\ \eta_5\\ \eta_6\\ \eta_7\\ \eta_8\end{bmatrix}$$

$$= \begin{bmatrix} p_0\eta_1 + p_1\eta_2 + p_2\eta_3 + p_3\eta_4 \\ q_0\eta_1 + q_1\eta_2 + q_2\eta_3 + q_3\eta_4 \\ p_0\eta_3 + p_1\eta_4 + p_2\eta_5 + p_3\eta_6 \\ q_0\eta_3 + q_1\eta_4 + q_2\eta_5 + q_3\eta_6 \\ p_0\eta_5 + p_1\eta_6 + p_2\eta_7 + p_3\eta_8 \\ q_0\eta_5 + q_1\eta_6 + q_2\eta_7 + q_3\eta_8 \\ p_0\eta_7 + p_1\eta_8 + p_2\eta_1 + p_3\eta_2 \\ q_0\eta_7 + q_1\eta_8 + q_2\eta_1 + q_3\eta_2 \end{bmatrix} \tag{A4}$$

Where p and q are determined as follows,

$$\begin{aligned}\left(C_n^{[4]}\right)^T C_n^{[4]} &= I_n \\ p_0 + p_1 + p_2 + p_3 &= \sqrt{2} \\ q_0 &= p_3 \\ q_1 &= -p_2 \\ q_2 &= p_1 \\ q_3 &= -p_0 \\ 0^0 q_0 + 1^0 q_1 + 2^0 q_2 + 3^0 q_3 &= 0 \\ 0^1 q_0 + 1^1 q_1 + 2^1 q_2 + 3^1 q_3 &= 0\end{aligned} \tag{A5}$$

In accordance with this manner, support length of "sup" of wavelet transformation matrix as well as any data size of wavelet transformation matrix can be determined as follows,

$$\begin{aligned}\left(C_n^{[\sup]}\right)^T C_n^{[\sup]} &= I_n \\ \sum_{j=0}^{\sup-1} p_j &= \sqrt{2} \\ q_j &= (-1)^j p_{((\sup-1)-j)} \quad (j = 0,1,2,...,(\sup-1)) \\ \sum_{j=0}^{\sup-1} j^r q_j &= 0 \quad \left(r = 0,1,2,...,(\frac{\sup}{2}-1)\right)\end{aligned} \tag{A6}$$

ACKNOWLEDGMENT

The author would like to thank Dr. Kaname Seto for his effort to conduct the experiments of data hiding.

REFERENCES

[1] K.Arai and K.Seto, Data hiding method based on wavelet Multi Resolution Analysis: MRA, Journal of Visualization Society of Japan, Vol.22, Suppl.No.1, 229-232, 2002.

[2] K.Arai and K.Seto, Data hiding method based on wavelet Multi Resolution Analysis: MRA utilizing eigen value decomposition, Journal of Visualization Society of Japan, Vol.23, No.8, pp.72-79,2003.

[3] K.Arai and K.Seto, Data hiding method based on wavelet Multi Resolution Analysis: MRA utilizing image space coordinate conversion, Journal of Visualization Society of Japan, 25, Suppl.No.1, 55-58,(2005)

[4] K.Arai, Keyword extraction method from paper based documents, drawings, based on knowledge importance evaluation with Analitic Hiearchical Process: AHP method, Journal of Image and Electronic Engineering Society of Japan, 34, 5, 636-644, (2005)

[5] K.Arai and K.Seto, Data hiding method based on wavelet Multi Resolution Analysis: MRA utilizing random scanning for improving invisibility of the secret image, Journal of Visualization Society of Japan, 29, Suppl.1, 167-170, 2009.

AUTHORS 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 1974 to 1978 also was with National Space Development Agency of Japan (current JAXA) from 1979 to 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 was appointed professor at Department of Information Science, Saga University in 1990. He was appointed 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 also appointed councilor of Saga University from 2002 and 2003 followed by an executive councilor of the Remote Sensing Society of Japan for 2003 to 2005. He is an adjunct professor of University of Arizona, USA since 1998. He also was appointed vice chairman of the Commission "A" of ICSU/COSPAR in 2008. He wrote 30 books and published 332 journal papers.