Chinese Geographical Science 2006 16(3) 276–281
DOI 10.1007/s11769-006-0276-y
www.springerlink.com

Digital Watermark-based Security Technology for Geo-spatial Graphics Data

JIA Peihong[1, 2], CHEN Yunzhen[1], MA Jinsong[1], ZHU Dakui[1]
(*1. The Key Laboratory of Coast & Inland Development of Ministry of Education, Nanjing University, Nanjing 210093, China; 2. Jiangsu Provincial Survey Bureau, Nanjing 210013, China*)

Abstract: The paper presents a set of techniques of digital watermarking by which copyright and user rights messages are hidden into geo-spatial graphics data, as well as techniques of compressing and encrypting the watermarked geo-spatial graphics data. The technology aims at tracing and resisting the illegal distribution and duplication of the geo-spatial graphics data product, so as to effectively protect the data producer's rights as well as to facilitate the secure sharing of geo-spatial graphics data. So far in the GIS field throughout the world, few researches have been made on digital watermarking. The research is a novel exploration both in the field of security management of geo-spatial graphics data and in the applications of digital watermarking technique. An application software employing the proposed technology has been developed. A number of experimental tests on the 1∶500,000 digital bathymetric chart of the South China Sea and 1∶10,000 digital topographic map of Jiangsu Province have been conducted to verify the feasibility of the proposed technology.

Keywords: geo-spatial graphics data; copyright protection; digital watermarking; stego carrier; data encrypting

1 Introduction

Geo-spatial graphics data, as fundamental data in building the Digital Earth, their security poses a big problem and draw much attention of many researchers all over the world.

Digital Watermarking, which is based on theories of steganography, is a novel technique for data security developed in the recent years. A digital watermark is defined as an identifiable digital signal or mode that is permanently embedded into other data, namely host data, while it does not affect the host data's usability (Ahmed and Day, 2004). So far researchers have come up with a variety of watermarking techniques which offer copyright protection for multimedia messages, such as image, video, audio, etc. However, very few researchers have engaged in the digital watermarking technology specifically designed for geo-spatial graphics data security. The research has developed a novel application of digital watermarking, and it should be a pioneer in the development of security management of geo-spatial graphics data.

Digital watermarking means to hide secret messages, including copyright protection and digital authentication messages, into geo-spatial graphics data. So it has been proposed as an efficient and effective way to identify legitimate owners and users. Digital watermarking also makes it technically possible to trace and proof the unauthorized distribution and reproduction of watermarked geo-spatial graphics data. Thus, illegal copiers and distributors could be captured more easily and the owner's legitimate interests could be further protected.

2 Principle and Feasibility

2.1 Principle

The principle of digital watermarking technique for geo-spatial graphics data is hiding a secret watermark with invisible copyright and user rights messages, or "watermark" (M), into an ordinary geo-spatial graphics data document named the cover carrier (C) so as to guarantee the security of geo-spatial graphics data by means of technology indirectly. The watermarked carrier is called the stego carrier (S). Fig. 1 shows the principle of hiding a digital watermark into geo-spatial graphics data carrier (Eggers and Girod, 2001; Jia, 2004, Lu, 2005).

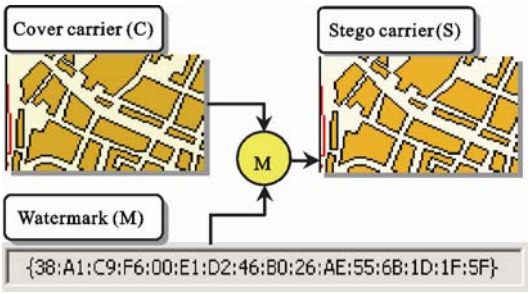

Fig. 1 Principle of hiding a digital watermark into GIS spatial data carrier

Received date: 2006-05-08; accepted date: 2006-08-15
Foundation item: Under the auspices of Jiangsu Provincial Science and Technology Fundation of Surveying and Mapping (No. 200416)
Biography: JIA Peihong (1972–), female, a native of Zhongwei of Ningxia Hui Autonomous Region, engineer, Ph.D. candidate, specialized in marine GIS theory and application. E-mail: jiapeihong@yahoo.com.cn

2.2 Fundamental requirements

2.2.1 Security

A digital watermark must be encrypted into a secret key to guarantee its security. As all message encrypting techniques, instead of protecting information directly, the watermark-based information hiding technology for geo-spatial graphics data aims at protecting the secret key of its watermark. Therefore all fundamental requirements of key in cryptology are equally applicable to the watermark-based information hiding technology for geo-spatial graphics data, for example, a large enough key-space (the key length in bits). In order to design a secure information hiding system, all issues concerning the generation, distribution and management of the secret key must be carefully, systematically, synthetically analyzed and addressed (Xenos, 2005）.

2.2.2 Imperceptibility

Imperceptibility means that the spatial data incur no perceptible visual distortions and measurement distortions after being embedded the watermark, namely the stego carrier (S) should be sufficiently identical with the original carrier (C). For example, if C is a large-scale map, then in a certain range of spatial resolution, the graphics of S and C can not be distinguished by either naked eyes or the computer (Su and Eggers, 2001).

2.2.3 Robustness

Robustness indicates the ability of a watermark to survive after its stego carrier is altered (Liu, 2002). For geo-spatial graphics data, the most severe alterations arise from data transformations, for example, as shown in Fig. 2, the transformation between E00 format and Coverage format in ArcInfo, or that between DWG/DXF format in AutoCAD and TAB/MIF/MID format in MapInfo. After the watermarked data undergo a transformation, watermark (M) could still be extracted from the altered data. Besides, the embedded watermark is of self-resilience, namely when the original data are destroyed or cropped by certain manipulation or conversion, the identifiable watermark could still be recovered and extracted from the data segments remained (Su and Eggers, 2001; Wu, 2005).

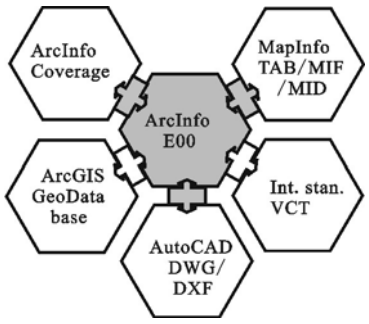

Fig. 2 Normal data transformations of watermark in GIS data

2.2.4 Capacity

Capacity means the amount of information that can be reliably hidden into the geo-spatial graphics data carrier. Ideally, if geo-spatial graphics data stego carrier would be assumed unaltered for ever, any amount of watermark information could be hidden into it without being perceived. However the assumption is unpractical, the capacity of a watermark is undoubtedly limited. As long as the watermark is imperceptible, the larger the watermark is, the less robust it is. Hence, a specific watermark embedding technique for geo-spatial graphics data involves the compromise among imperceptibility, robustness and capacity (Shih, 2003).

2.3 Feasibility

The feasibility of hiding digital watermark into geo-spatial graphics data can be argued with the following two points. First, the formats now wildly used in geo-spatial graphics data are all redundant in bits. So it is technically feasible to hide a digital watermark into geo-spatial graphics data without impairing the data's precision or undermining their distribution and usability. Second, the masking effect of human visual system (HVS) reduces the sensitivity of HVS to minor modifications in geo-spatial graphics (Xenos, 2005). Therefore it is possible to take advantage of this effect to invisibly insert a watermark into the geo-spatial graphics data carrier.

3 Key Techniques

Hiding digital watermark into geo-spatial graphics data is composed of three major components: watermark generation, watermark embedding, and watermark extraction/detection (Lin, 2000). Watermark extraction/detection techniques are simpler. So in this section, we present in detail the techniques of watermark generation and embedding. Other key techniques concerning watermark-based information hiding include watermark messages encrypting, as well as watermarked geo-spatial graphics data compressing and encrypting.

3.1 Watermark generation

In the research, the copyright message and user ID are converted into a unique user sequence number that in turn is encrypted into a watermark secret key. Then the secret key is inserted into geo-spatial graphics data. It is in this way that watermark messages are hidden into geo-spatial graphics data. Therefore, user sequence number generator is essentially watermark secret key generator.

Pseudo Random Number Generator is a kind of user sequence number generator. Pseudo Random Number Generator may generate 128 bits (16 bytes) binary number sequences based on a seed from the computer system timer. There are two functions from which Visual C++ generates a random number: rand (void) and srand (seed). The rand() function returns a pseudorandom integer in the range 0 to RAND_MAX. RAND_MAX is a constant (Declaration: #define RAND_MAX 0x7fff). The srand() will be called before the calls to rand() to sets the starting point for generating a serious of pseudo-random integers, e. g. srand ((unsigned) time

(NULL)). The random number sequences will be generated differently when a random number is initiated as a seed. The system timer on a computer will be set as a seed in this research. The calls to the function rand() usually have the different timer values. Consequently, this is the way to guarantee that any user sequence number in the geo-spatial graphics data is random and unique.

3.2 Watermark embedding

3.2.1 Mark-sequences locations

As presented in section 3.1, the watermark generated is of 128 bits (16 bytes). The 128-bit watermark consists of 16 mark-sequences in a specific order, and every mark-sequence consists of 8 bits as well. Therefore watermark embedding technique has to do with the determinations of mark-sequence locations as well as mark-bit locations.

Geo-spatial graphics data are a spatial information carrier, which is mainly composed of space coordinate series files. Taking E00 files in ArcInfo for example, the space coordinates in Arc section are classified into two categories, one is embedded with mark-sequences, and the other is not. As the watermark embedding positions must be distinguishing from one E00 files to another, the statistical characteristics of space coordinates must be computed to determine the mark-sequences locations. The proposed technique calculates the mean distances of all space coordinates, picks out 16 space coordinates with the least mean distances as marked locations of watermark embedding, and arranges the 16 locations in order as their mean distances increase. The embedding order of the 16 mark-sequences of the watermark secret key is thus determined. Since the length of marked locations is smaller than that of the total data, the watermark data can be efficiently hidden. In the process of watermark extraction, the 16 marked locations hiding watermark can be determined likewise, namely by calculating the mean distances of all space coordinates. After the specific data of mark-sequences are retrieved, rearrange the 16 sequences in order as their corresponding locations' mean distances increase.

Then pseudo noise code generator is employed to decide which data in space coordinates are to be watermarked. Fig. 3 shows the scheme of computing mark-sequences embedding locations. The blue curve is noise frequency curve, or the randomly generated pseudo noise code. Watermark data are then added into the curve according to its statistical character. The red points are the 16 data points with the least deviations from the mean of noise frequency curve. Consequently the 16 red points are signed as singular points where the 16 mark-sequences of watermark are exactly embedded (Lin, 2000).

3.2.2 Mark-bits locations

Mark-bits embedding and extraction in geo-spatial graphics data can employ neither routine spread spectrum scheme, nor random interval watermarking or pseudo

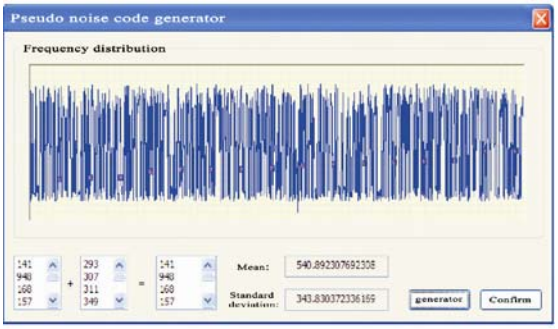

Fig. 3 Pseudo noise code generator for determining the places to be watermarked

random permutation scheme. This is because the above three schemes all be applied to multimedia data in streaming format , while spatial data are of non-linear spatial topological relationship, so it is impossible to arrange spatial data into the same linear sequences in different GIS software, nor to store space coordinates in a streaming order. Therefore some specific digital watermarking techniques are required in order that mark-bits embedding positions will not change with data editing manipulation and data transformation.

For purpose of mark-bits locations keeping stable after the general GIS spatial data transformations, the storage precisions of common GIS data formats are investigated, as shown in Table 1. It is observed that mark-bits can be located into the general significant bits of GIS spatial data without impairing the data's precision.

Table 1 Storage precision of common GIS data

Data format	Significant data precision
E00/ Coverage in ArcInfo	10–15digitals（double）
TAB/MIF/MID in MapInfo	8 digitals
DWG/DXF in AutoCAD	6–16digitals (optional)

The proposed technique employs LSB algorithm (Least Significant Bits algorithm) to embedding mark-bits. LSB is the most typical watermarking algorithm which takes a spatial-domain approach. The embedding process consists of two steps. First, randomly pick out a subset of spatial data carrier elements. Second, conduct substitution operations, to replace carrier elements' LSB with mark-bits. For example, as for E00 spatial data files in ArcInfo, every carrier element's two least significant bits are replaced with two mark-bits. In the reverse extraction process, the stego carrier elements' LSB are extracted and arranged in certain order to reconstruct 16 mark-sequences one by one (Lin, 2000).

The advantages of LSB, as a hiding information algorithm for image data, are firstly that a larger amount of watermark information can be hidden; secondly, that imperceptibility can be guaranteed. However, LSB is vulnerable to filtering, quantization, geometric distortion and other manipulations. The two most distinctive attributes of geo-spatial graphics data are the topological

relationship record and the accuracy of spatial position. Consequently, in the research a hybrid technique is employed to make full use of the advantages of LSB, while at the same time, to avoid its weakness in robustness. The hybrid technique is composed of LSB combined with the data's topological relationship record, as well as watermark and data package encrypting. Besides, since in this research the watermarks embedded into geo-spatial graphics data are for purpose of tracing secretly, not all geo-spatial graphics data are watermarked, but some dozens of statistically singular points in the whole graphics are assigned as watermarking locations. Further, the digital watermarks are randomly scattered into geo-spatial graphics data using pseudo noise code generator. Thus watermark data are very hard to attack due to the irregularity of watermarking locations. Confronted with the hybrid technique, the attacker will not know whether later mark-sequences are inserted in the previous order, nor will he discover the exact LSB hiding mark-bits. So no watermarking regularity will be detected. In addition, the watermark and data package are both encrypted. If an illegal user attempts to remove or bypass the watermark, or takes the risk of attacking maliciously, the data precision will be totally degraded and the data useless. In a word, the hybrid LSB approach designed for geo-spatial graphics data possesses a better robustness.

3.3 Watermark and geo-spatial data encrypting

Digital watermarking combined with data encrypting is a significant way to defend the watermark against aggressive attackers and enhance the security of watermark. The proposed technology uses data encryption twice. The first encryption is encoding digital watermark (M) before it is hidden into geo-spatial graphics data cover carrier (C). Thus even the illegal user is aware of the existence of hidden watermark, it is very difficult for him to discover it, and of course more difficult to decode it. The second one is encrypting watermarked geo-spatial graphics data files for purpose of enhancing security of data distribution.

There are many key cryptosystems available for watermark and geo-spatial graphics data encrypting, for instance, the symmetric key cryptosystem DES (Data Encryption Standard) and IDEA (International Data Encryption Algorithm). In the research DES is chosen as the key cryptosystem and its experimental results are favorable.

DES is a 128-bit double encrypting algorithm, developed by IBM in 1997. Due to its high speeds in encrypting and decrypting, as well as its distinguished security, DES has long been the most widely used cipher in commercial and financial applications. As the Federal Information Processing Standard published by the U.S. government, DES is possessed of a prominent position in the field of cryptology. DES is a type of iterated cipher. The 64-bit input plaintext block is first divided into two 32-bit blocks. Then there are 16 rounds of transformations. In each round, the two blocks are both subjected to a complex key-dependent computation performing shift and substitution operations. Finally the 64 bits are scattered and transposed to arrange a new output ciphertext block which is so secure that there exists no better known attack other than exhaustive key search. Key, Data, Mode are three entry 8-byte (64-bit) parameters of DES. Key is the secret key of DES algorithm, Data is the would-be encrypted or decrypted data, and Mode indicates to put through an encryption or decryption (Wu and Chang, 2005).

3.4 Watermarked data compressing

The watermarked and encrypted geo-spatial graphics data have be packaged and compressed to reduce the data volume no matter they are communicated via CD, the Internet, or any other medium. Data compressing not only facilitates the data communication, but also provides another counter measure for the data.

LZW (Lossless Compression Algorithm) is a common algorithm processing geo-spatial graphics data with high fidelity. Since LZW compresses the data storage space, rather than remove the data bits, the stego data's LSB (Least Significant Bits) will not be destroyed, thus the embedded watermark will not be impaired. Experimental results show geo-spatial graphics data can be reduced by more than 20% in volume after being compressed by LZW.

4 Technology Implementation

The hybrid of digital watermarking, data encrypting and compressing technique guarantees the secure communication and distribution of geo-spatial graphics data. The security precautions will be relieved when the data are installed by an authorized user. The flowchart of implementation of watermark-based information hiding technology for geo-spatial graphics data and its inverse is shown in Fig. 4.

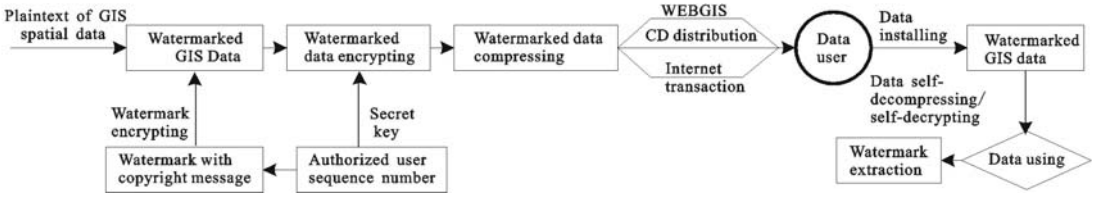

Fig. 4 Flow chart of watermark-based information hiding

As shown in Fig. 4, the operations of watermarking, encryption and compression are carried out for the data before they are used. So the unauthorized data obtained through the illegal channel can not be used because they can not be installed in the computer. Since data installation is companied by data decompression and decryption, the data will be installed only if the authorized user ID is provided as well as copyrights and user rights agreement are accepted. The installed data restore to watermarked plaintext files, and then they can be used. At the same time, the digital watermark starts to carry out its penetrating permanent tracing mission.

5 Experimental Results

5.1 Watermark extraction/detection

Watermark extraction/detection techniques available are simple. Direct detection and correlation detection are two common techniques, while the proposed technology employs Maximum A Posterior-based detection. Many tests are conducted to verify that the watermarks can be successfully extracted and detected. For example, Fig. 5 shows a data producer's user record management database. Fig. 6 shows a 1∶10,000 topographic map of Jiangsu Province that has been watermarked. Its stego geo-spatial graphics and cover graphics overlap totally. A watermark secret key and relevant user messages are detected from the stego geo-spatial graphics, which are identical with the corresponding key and messages listed in the user management database. The existence of watermark in the coverage is thus successful verified.

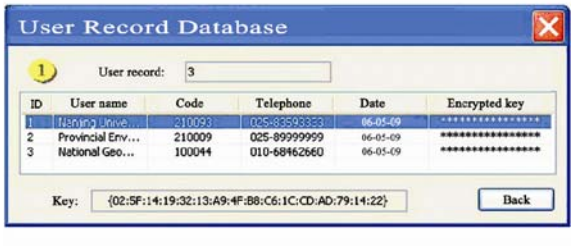

Fig. 5 Data user record management database

Fig. 6 Detecting watermark messages from watermarked data

5.2 Watermark imperceptibility and data accuracy

In order to evaluate the performance of watermark imperceptibility and data accuracy, the watermarked stego geo-spatial graphics is overlaid onto the original graphics to measure its error. The error is 0.03m, which is much less than 1m (the Chinese accuracy criterion). Furthermore, when the watermarked geo-spatial graphic is magnified 100 times, no visual distortion and accuracy distortion are perceived by naked eyes. Therefore the stego geo-spatial graphics data are sufficiently identical with the original one.

Nevertheless there still exists error arising from hiding watermark into geo-spatial graphics data, though it can meet the accuracy requirement. Hence is it the case that the original overlapped nodes become unclosed after watermarking? A procedure is designed to judge whether a singular point where mark-sequence is embedded is a node. If it is a node, then we pick out all nodes overlapped with it and mark all of them. Thus the original nodes maintain overlapped after being watermarked. This has been tested by optionally magnifying the watermarked geo-spatial graphics. Thus watermark imperceptibility and data accuracy are both verified.

5.3 Watermark robustness

A robust watermark is resistant to a variety of disturbances. The expected disturbances arise from all kinds of data manipulations, including routine secure manipulations and malicious attacks (Sun and Sun, 2004). Data transformation is a common non-malicious disturbance. Now there is a multitude of GIS software, and there is no uniform GIS software platform. Therefore, the producer of geo-spatial graphics data has to deal with data transformation in data publishing, and the user will also transform data from a format to another in developing his own thematic application system. All file formats of GIS software available in the market now are not open to the public. As a result, though all software provide the data transformation interface, it is impossible that the file formats of other GIS software could be thoroughly figured out, thus the data transformation is incomplete, namely that not all data messages are transformed correspondingly. This issue has long perplexed the data producer during the process of data management and publishing. Hence, the robustness against a variety of data transformations is an important performance indication of digital watermarking in the research.

In the tests, the watermarked E00 data in ArcInfo are transformed into other common GIS graphics data formats, such as Coverage and Sharp in ArcInfo, DWG/DXF in AutoCAD, and TAB/MIF/MID in MapInfo. Then when the data are retransformed into E00 format, the watermark are successfully extracted and detected.

Can the watermark survive the normal geo-spatial graphics data editing? Is there any difference when the proposed technology is applied to a binary system and a digital one? In order to settle the two problems, the user's

demand of geo-spatial graphics data is analyzed. Generally, geo-spatial graphics data are used to create a base map which is in turn added with thematic data. Consequently, the data user relies much on the accuracy of geo-spatial graphics data, and he will not modify the data rashly. Further, a number of tests on the 1∶500,000 digital bathymetric chart of South China Sea and 1∶10,000 digital topographic map of Jiangsu Province are conducted. It is verified that routine manipulations to the stego data in ArcInfo software will not affect the detection of watermark, nor will the data transformation from E00 digit format into Coverage and Sharp binary format.

Many experiments are also conducted to test the robustness performance against malicious attacks. It is verified that the embedded watermark can not be removed or decrypted without rendering the attacked data inaccurate as long as the watermarking algorithm is unknown. The inaccurate geo-spatial graphics data are totally useless even they are stolen.

6 Conclusions

The application software of the proposed technology has been developed with VC++ language, a number of tests have been repeatedly conducted on the 1∶500,000 digital bathymetric chart of South China Sea and 1∶10,000 digital topographic map of Jiangsu Province and the feasibility of proposed technology has been verified. The presented study will provide law enforcement officials with technical support for tracing and identifying piracy whereby unauthorized reproduction of basic geo-spatial data can be effectively banned and the market of survey and map products can be fully regulated as well. In the same way, IPR (Intellectual Property Rights) laws and regulations can be more strictly enforced so as to effectively punish the divulgence and piracy as well as to improve the security of GIS geo-spatial data being traded and published via the Internet.

Some technical details of the proposed technology still need to be further tested and improved in practice so that the technology would be more robust, practical and sophisticated for the data security management of the Digital Sea and the Digital City projects. In addition, the watermark embedding and encrypting techniques should be further updated and improved to be better integrated into geo-spatial graphics data.

References

Ahmed Ayman M, DAY Dwight D, 2004. Applications of the naturalness preserving transform to image watermarking and data hiding. *Digital Signal Processing*, 14: 531–549. DOI:10.1016/j.dsp.2004.08.002

Eggers Joachim J, Girod Bernd, 2001. Quantization effects on digital watermarks. *Signal Precessing* , 81: 239–263.

Jia Peihong, 2004. Technical methods for encrypting and hiding digital watermark in GIS spatial data. *Geomatics and Information Science of Wuhan University*, 29(8): 747–750. (in Chinese)

Lin Phenlan, 2000. Robust transparent image watermarking system with spatial mechanisms. *The Journal of Systems and Software*, 50: 107–116.

Liu Jianchyn, Chen Shuyuan, 2001. Fast two-layer image watermarking without referring to the original image and watermark. *Image and Vision Computing*, 19: 1083–097.

Lu Chunshien, 2005. Towards robust image watermarking: combining content-dependent key, moment normalization, and side-informed embedding. *Signal Processing: Image Communication*, 20: 129–150. DOI:10.1016/j.image.2004.10.002

Shih Frank Y, Wu Scott Y T 2003. Combinational image watermarking in the spatial and frequency domains. *Pattern Recognition*, 36: 969–975.

Su Jonathan K, Eggers Joachim J, 2001. Analysis of digital watermarks subjected to optimum linear filtering and additive noise. *Signal Processing*, 81: 1141–1175.

Sun Rui, Sun Hong, 2004. Application of transmit diversity for improved robust watermarking. *Journal of Systems Engineering and Electronics*, 15(2): 163–170.

Wu Hsienchu, Chang Chinchen, 2005. A novel digital image watermarking scheme based on the vector quantization technique. *Computers & Security*, 24: 460-471. DOI:10.1016/j.cose.2005.05.001

Wu Jsienchu, 2005. A novel digital image watermarking scheme based on the vector quantization technique. *Computers & Security*, 24: 460–471.

Xenos Michalis, 2005. A model for the assessment of watermark quality with regard to fidelity. *J. Vis. Commun. Image R.*, 16: 621–642.