SPATIAL-SPECTRAL WATERMARKING SCHEME FOR JPEG STEGANOGRAPHY

B. Marques, B. Sim and L. Cheng *

** School of Electrical and Information Engineering, University of the Witwatersrand, Private Bag 3, Wits. 2050, Johannesburg, South Africa.*
Emails: bradley.marques@students.wits.ac.za; benjamin.sim.za@ieee.org; ling.cheng@wits.ac.za

Abstract: Steganography is the art of concealing the presence of a communication, across an open channel, from undesired parties. This is achieved by the encoding of message data in innocuous cover objects, such as JPEG digital images. Many JPEG steganography algorithms exist; most perform the encoding in the spectral domain of the image. A new steganographic method, deemed the Spatial-Spectral Watermarking Scheme (SSWS), has been developed and utilises both the spatial and the spectral domains for the encoding process. A developed mathematical model of the scheme predicts a maximum gain in terms of steganographic capacity of 4.9 over an existing algorithm, JSteg. Experimental data obtained from a software prototype confirms this result. Furthermore, the experimental data confirms that the securities in the spatial domain and the spectral domain are inversely proportional to the information amounts embedded in these two domains.

Key words: Data hiding, Information theory, JPEG, Steganography, Steganalysis, Watermarking.

1. INTRODUCTION

Steganography is a field of information theory involving the concealment of the presence – not merely the content – of communication between two parties from an outside observer [1–5]. In modern digital steganography, this is achieved by hiding the message data within innocuous cover media such as image, audio or video files. The Joint Photographic Experts Group (JPEG) digital image format has become a *de facto* standard in the field of digital photography and is prevalent on the World Wide Web [6]. It thus offers a perfect platform for digital image steganography. The applications of JPEG steganographic systems include covert communications, embedding of image metadata within the image itself, and digital image watermarking for copyright protection.

Existing JPEG steganographic schemes – such as JSteg, OutGuess and J3 – utilise the spectral domain of the JPEG file to conceal secret message information [7, 8]. A new method of JPEG steganography, the Spatial-Spectral Watermarking Scheme (SSWS), which utilises both the spectral and the spatial domains of a JPEG image has been designed, developed and investigated. The linking of the two domains is made possible through the use of Forward Error Correction (FEC), allowing errors introduced by the dual encoding to be recovered. By utilising both domains of the image, the capacity and security of the encoding process is increased.

The remainder of this paper is organised as follows: relevant background information on JPEG steganography is presented in Section 2. in terms of steganographic concepts, existing JPEG steganographic systems and steganalysis. Concepts of FEC utilised in SSWS are also highlighted. The design of the SSWS is then examined in Section 3. A mathematical model describing the SSWS is presented in Section 4. Experimental data collected to verify this model in terms of capacity and security are also presented in this section. Finally, a conclusion to the paper is presented in Section 5.

2. PRELIMINARIES

2.1 Steganography and Steganalysis

Steganography is the art of hiding a secret message in an innocuous medium, deemed the *cover* object. Steganography focuses on the concealment of the presence of communication - not necessarily the content thereof, as is the case with cryptography [1,3,5]. The message data are embedded into the cover to produce the *stego* object [3].

The strength of digital image steganographic systems is defined by three main criteria: capacity, security and robustness. Capacity is a measure of the amount of information encodable within a cover image. Security represents an algorithm's resistance to a steganalytic attack [1, 3, 9]. Such an attack may be targeted or blind: targeted techniques use method-specific limitations or features that are introduced during encoding to detect the presence of a message: blind, or universal, techniques use statistical attacks, such as the chi-squared test, to detect message presence [1, 2, 10]. Other techniques, deemed quantitative staganalysers, use statistical methods not only to detect the presence of the message but to also estimate its length [7, 9]. Steganographic robustness defines a technique's ability to recover a message after tampering. Such tampering may be active – intentionally altering the content of an embedded message to disrupt the steganography – or passive – such as cropping or blurring an image [11]. However, robustness is not a factor of the encoding algorithm alone, but is also dependant on any FEC techniques applied to the data prior to embedding. Thus, it will not be considered further in this paper.

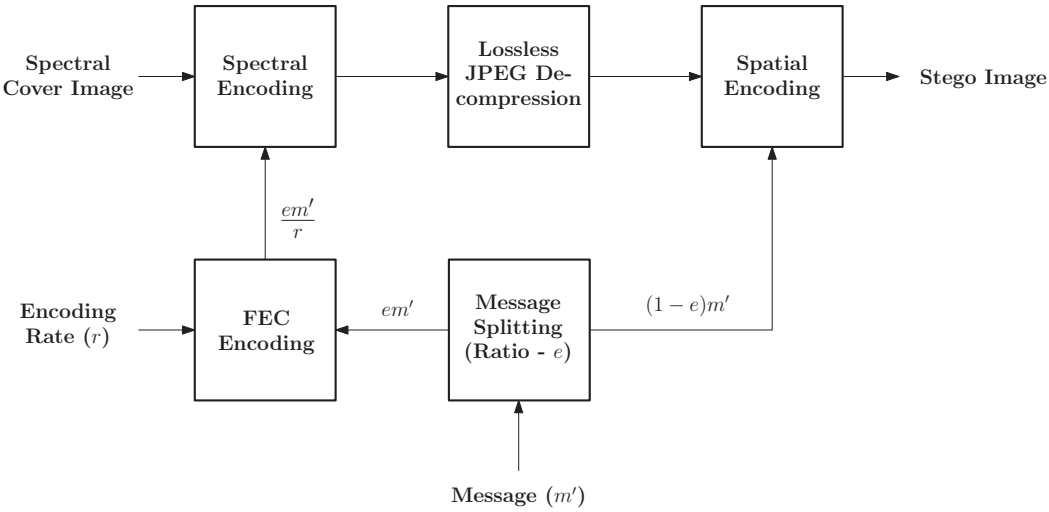

Figure 1: The encoding process for SSWS

2.2 *Existing Steganographic Techniques*

Existing JPEG steganographic techniques typically perform the encoding process in the frequency domain of the image [3]. Thus, changes are elicited in the spectral domain – or the Discrete Cosine Tranform (DCT) coefficients – of the JEPG image. Additionally, many techniques utilise the concept of Least Significant Bit (LSB) encoding [2, 3]. This minimises the change to the image, since only the selected DCT coefficient's LSB is altered, not the entire value. For instance, the JSteg algorithm sequentially replaces each DCT coefficient's LSB with the required message bit [2, 3]. The OutGuess 1.0 algorithm relies on a similar scheme as JSteg, but the selection of DCT coefficients in which to embed is more advanced; the algorithm performs a pseudorandom walk over the entire image spectral domain, embedding the message at the visited locations [3, 7, 12]. The seed of the pseudorandom number generator thus becomes a private key for the communication [13]. To increase algorithm security against statistical blind steganalysis, methods such as OutGuess 2.0 and J3 attempt to maintain histogram neutrality during the embedding process [3, 8]. This is achieved by introducing counteractive changes for every change that is introduced during embedding.

2.3 *Forward Error Correction*

FEC is a type of error correcting code used in the field of telecommunications, and involves the addition of procedurally generated redundant information to a message for the purpose of noise and data loss insensitivity [14]. The redundancy is procedurally generated from the message data and enables the receiver to recover the sent message in the event of induced errors or erasures in the received message [14, 15]. Various FEC techniques exist which offer varying levels or error and erasure correcting capabilities.

3. THE SPATIAL-SPECTRAL WATERMARKING SCHEME

The existing algorithms discussed in Section 2.2 define the actual process of data embedding. The developed SSWS, however, does not. Rather, it defines a scheme to improve the performance of any existing steganographic algorithm. SSWS defines a protocol for the encoding process whereby message data are encoded within both the spectral and the spatial domains of the image. Any existing or new technique may be used as an underlying algorithm.

The SSWS encoding process, illustrated in Figure 1, initially interleaves the message. The interleaved message is divided into two sections, m_1 and m_2 defined by a splitting factor, $e \in [0, 1]$.* Since the SSWS embeds in both spatial and spectral domains, FEC encoding is required to facilitate error recovery. The m_1 message section undergoes FEC encoding, which adds parity information to facilitate error recovery in the event of the spatial embedding introducing errors into the spectral message section. The FEC technique used was a block code with the code rate r. The encoded m_1 is then embedded (by the underlying algorithm) within the spectral domain. The spectral cover image is losslessly decompressed to generate a spatial image. The remainder of the message, the m_2 message section, is then embedded into this spatial image by the same underlying steganographic algorithm.

By utilising both domains, the SSWS can offer improved capacity and security (see Section 4.). The capacity is increased as it uses two separate domains into which the information can be embedded. Furthermore, SSWS offers a two-fold improvement to security. Firstly, by splitting the message content between two domains, the message content (in bits) is reduced within each domain for a given payload. Thus, steganalytic attacks targeting either domain have less information to detect. Secondly,

* e indicates the percentage of the message embedded within the spectral domain

by performing the double encoding, the first layer of information is effectively obscured by the second encoding process. Thereby, improving the systems resistance to spectral based steganalysis techniques. Note that this scheme can be easily modified and applied to a McEliece public-key cryptosystem, thanks to the FEC process used to bridge between the two domains.

4. RESULTS AND ANALYSIS

4.1 Mathematical Model of SSWS Capacity

In order to develop a mathematical model of the SSDDW scheme, it is necessary to define the capacity gain, g, that the technique offers over its underlying algorithm as

$$g = \frac{m' - m}{m}, \tag{1}$$

where m' is the maximum length of embeddable message by the SSWS, and m is the maximum length of embeddable message using the standard algorithm. The message m' is split by the ratio e into two smaller messages, as shown in Figure 1. As mentioned in Section 3., the component embedded in the spectral domain undergoes FEC encoding with a code of rate r. The two message sections, m_1 and m_2 are thus written as

$$m_1 = \frac{em'}{r} \tag{2}$$

and

$$m_2 = m'(1-e) \tag{3}$$

Two independent constraints are present in the system: the capacity of the spatial domain and the capacity of the spectral domain. It is evident that for large images, using LSB embedding techniques, a maximum of one bit per image pixel can be embedded in the spatial domain. Hence,

$$m'(1-e) \leq p, \tag{4}$$

where p is the number of pixels in an image (including all three colour channels). Similarly, it is known that for the standard algorithm, the maximum embeddable message m is

$$m = p\alpha. \tag{5}$$

From experimental results, it was found that in the spectral domain there is a linear relationship between image size, p, and algorithm capacity. This was found to hold true for all investigated embedding techniques, as may be seen in

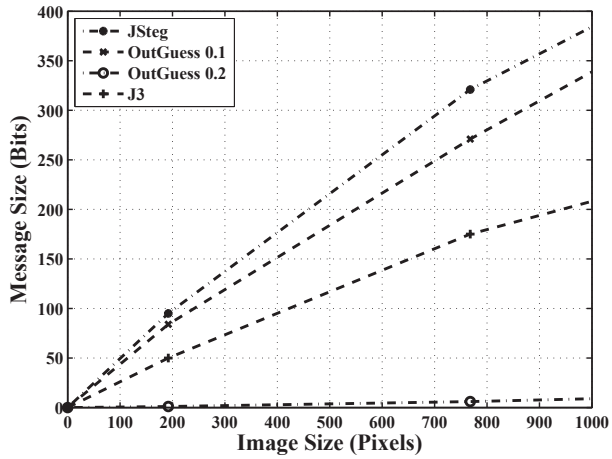

Figure 2: The measured linear relationship between image size and message bits embedded in the spectral domain within an image.

Figure 2. The relationship between p and spectral capacity is thus a constant value, α (in bits per pixel) for a given embedding algorithm. The constraint thus placed on the spectral domain capacity is:

$$\frac{em'}{r} \leq p\alpha. \tag{6}$$

By solving (1) for m', substituting into (4), and using (5), a maximum bound is found for g as follows

$$g \leq \frac{p}{\alpha(1-e)} - 1. \tag{7}$$

Similarly, by using (6), the following constraint is found

$$g \leq \frac{r-e}{e}. \tag{8}$$

Combining these two constraints in (7) and (8) leads to the upper bound of the capacity gain, g, of the SSWS system as

$$g \leq \min\left(\frac{p}{\alpha(1-e)} - 1, \frac{r-e}{e}\right). \tag{9}$$

As mentioned in Section 2.1, capacity is a defining characteristic of steganographic systems. The upper bound of (9), was used with a known α (for the JSteg technique) to produce the surface plot of Figure 3, which shows the theoretical maximum gain as a function of the message split percentage e and FEC code rate r. As may be seen, the estimated maximum gain is around 4.9 at a message split ratio of 0.16. This theoretical maximum gain is obtained with an FEC code rate of 1 (no parity information added). However, this is impractical since this theoretical maximum assumes that no errors are introduced by embedding in the spectral domains (see Section 3).

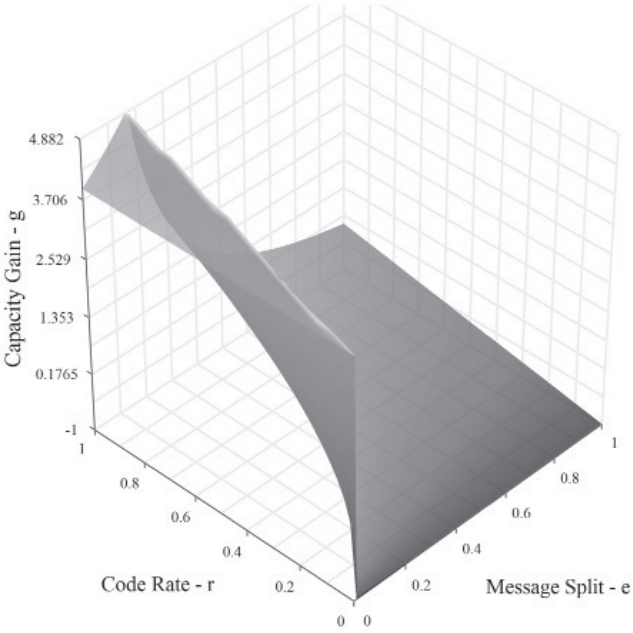

Figure 3: Predicted results for capacity gain

Experiments were conducted to validate the model. The maximum embeddable message length (capacity) was determined as a function of the message split ratio e for a given technique and code rate. The JSteg technique, which has an α value of 0.203 bits per pixel (refer to Figure 2) was used with a Reed-Solomon code of rate $\frac{4}{7}$. The empirical data collected in addition to the predicted bounds of the model (9), are shown in Figure 4. As may be seen, the experimental results are in accordance with the bounds of the theoretical model. As shown, with 7% of the message embedded in the spectral domain, the SSDDW scheme provides a maximum gain of 4.4 over the standard JSteg algorithm. This gain in capacity is offset by a decrease in algorithm security, which is examined in the next subsection.

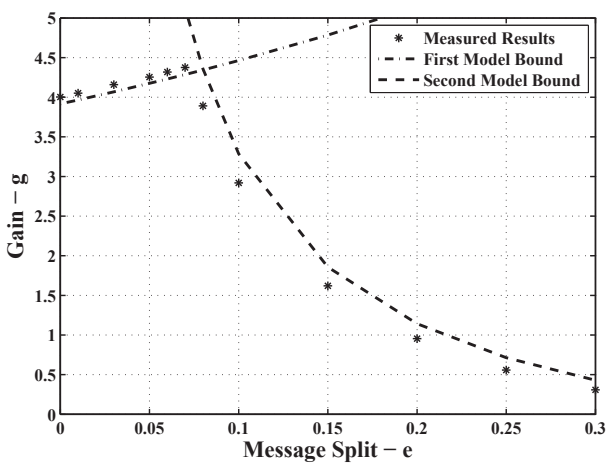

Figure 4: The gain in capacity of the SSWS system, using JSteg, for different message split ratios e. The theoretical bound detailed in Section 4.1 are shown to accurately predict the measured behaviour.

4.2 Security

As mentioned in Section 2.1, the security of embedding – the degree to which the technique resists steganalysis – is a defining characteristic of steganographic systems. A few measures, targeting both the spectral and spatial domains, were used to test the security of the SSWS. Since the method utilises both spectral and spatial domains, security must be tested in both of these domains.

The Kullback-Leibler (KL) distance is a metric for steganographic security which quantifies the difference between the cover and stego image histograms. It is thus a measure of the change induced in the spectral domain during message embedding. An experiment was performed in which the KL distance was measured for a stego image produced from embedding a message of fixed length (10 kbit) into a test cover image, under a varying percentage split e. The obtained KL distances are inverted (since lower KL distances indicate a more secure system) and are expressed as a percentage improvement over the inverse KL distance obtained with the standard underlying algorithm, JSteg. The experimental results obtained may be seen in Figure 5. As may be expected, the security of the technique decreases with increasing e. This is the case since, as more message is encoded into the spectral domain, steganalysis techniques targeting this domain are more likely to detect message presence.

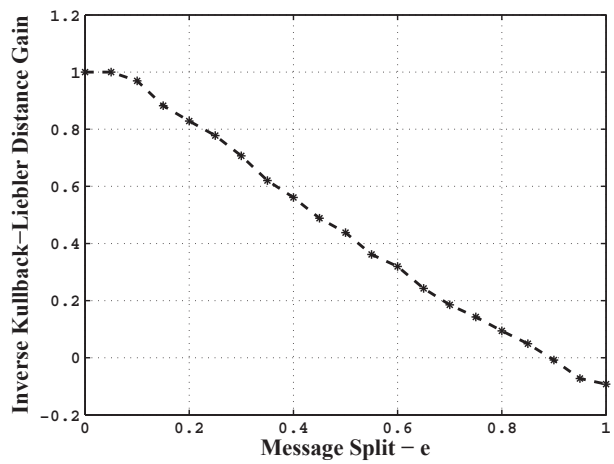

Figure 5: The gain in spectral security of the SSWS, using JSteg, for different message split ratios e.

The field of LSB steganalysis is a mature field, with algorithms with proven performance [16–18]. Thus, security testing in the spatial domain was performed using the Sample Pairs (SP) analysis as presented in [18]. A set message size and set message content was used for the experiment. The message split e was varied, and measurements of the estimated message length taken. These results were then compared to the message length estimated by the SP analysis for a stego image produced by JSteg under the same conditions. The relative gain in security afforded by the SSWS was thus determined. A linear regression trendline of the obtained data, as illustrated by Figure 6 was determined. As may be seen,

the SSWS has poor security in the spatial domain. The results indicate that the spatial security increases as more data is embedded in the spectral domain, as may be expected.

Thus, whilst the SSWS increases both the capacity and spectral domain security of this steganographic technique, the encoding of data in the spatial domain is highly vulnerable to steganalysis techniques such as the SP analysis. The choice of message splitting ratio *e* must thus depend on the application, and whether capacity or security is important.

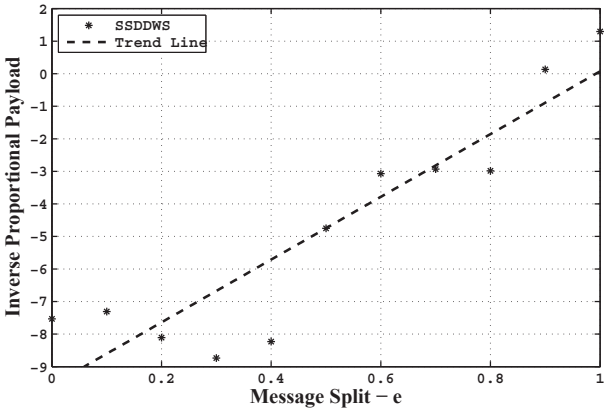

Figure 6: The gain in spatial security of the SSWS, using JSteg, for different message split ratios *e*.

5. CONCLUSION

Steganography is the concealment of secret information in innocuous cover media such that the very presence of the communication is hidden. This paper presents a new technique deemed the Spatial-Spectral Watermarking Scheme (SSWS) which uses a spatial-spectral embedding process to improve steganographic performance. The message is split across both the spatial and the spectral domains of the image and is embedded by an underlying stenographic algorithm. FEC is used to recover from errors introduced in the embedding process. The algorithm was tested in terms of its capacity and security relative to its underlying technique. Through the development of a mathematical model and experimental testing, the capacity gain afforded by use of the SSWS was shown to be 4.4 over the existing JSteg algorithm. This result was obtained with 7% of the message embedded in the spectral domain, and the remaining 93% embedded in the spatial domain. The splitting technique of SSWS improves the steganographic security since the experimental results (refer to Figure 5 and Figure 6) show that the securities in both domains are inversely proportional to the information amount embedded in either domain.

ACKNOWLEDGEMENT

The authors would like to acknowledge the CeTAS group and Prof Ivan Hofsajer of the University of the Witwatersrand as well as Phil Sallee [19], Rick Rosson [20] and Jessica Fridrich [21] for various MATLAB functions used in the prototype software.

REFERENCES

[1] T. Pevny and J. Fridrich, "Determining the stego algorithm for JPEG images," in *Information Security, IEE Proceedings*, vol. 153, no. 3. IET, 2006, pp. 77–86.

[2] J. Kodovsky and J. Fridrich, "Quantitative structural steganalysis of Jsteg," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 4, pp. 681–693, 2010.

[3] N. Provos and P. Honeyman, "Hide and seek: An introduction to steganography." IEEE Computer Society, 2004, pp. 32–44.

[4] C. Cachin, "An information-theoretic model for steganography," in *Information Hiding*. Springer, 1998, pp. 306–318.

[5] J. Fridrich, T. Pevnỳ, and J. Kodovskỳ, "Statistically Undetectable JPEG Steganography: Dead Ends Challenges, and Opportunities," in *Proceedings of the 9th workshop on Multimedia & security*. ACM, 2007, pp. 3–14.

[6] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, pp. 30–44, April 1991.

[7] J. Fridrich, M. Goljan, and D. Hogea, "Attacking the outguess," in *Proceedings of the ACM Workshop on Multimedia and Security*, 2002, pp. 967–982.

[8] M. Kumar and R. Newman, "J3: High payload histogram neutral JPEG steganography," in *IEEE 2010 Eighth Annual International Conference on Privacy, Security and Trust*, aug. 2010, pp. 46–53.

[9] T. Pevnỳ, J. Fridrich, and A. Ker, "From blind to quantitative steganalysis," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 445–454, April 2012.

[10] S. Agaian and P. Perez, "Block structure chi-(p,q) steganalysis method," in *Region 5 Conference, 2006 IEEE*, April 2006, pp. 307–311.

[11] C. Smith and S. Agaian, "Denoising and the active warden," in *IEEE International Conference on Systems, Man and Cybernetics*, oct. 2007, pp. 3317–3322.

[12] N. Provos, "Defending against statistical steganalysis," in *10th USENIX security symposium*, vol. 10. Washington DC, USA, 2001, pp. 323–336.

[13] N. Hopper, J. Langford, and L. von Ahn, "Provably secure steganography," in *Advances in Cryptology: CRYPTO 2002*, M. Yung, Ed. Springer Berlin / Heidelberg, 2002, vol. 2442, pp. 119–123.

[14] W. C. Huffman and V. Pless, Fundementals of Error Correcting Codes. Cambridge University Press, 2003.

[15] J. C. Moreira and P. G. Farrell, Essentials of Error-Control Coding. John Wiley & Sons Ltd, 2006.

[16] A. Ker and R. B¨ohme, "Revisiting weighted stego-image steganalysis," Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, vol. 6819, p. 5, 2008.

[17] A. Ker, "A general framework for structural steganalysis of LSB replacement," in Information Hiding. Springer, 2005, pp. 296–311.

[18] S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSB steganography via sample pair analysis," in Information Hiding. Springer, 2003, pp. 355–372.

[19] P. Sallee, "Matlab JPEG toolbox," http://www.philsallee.com/jpegtbx/index.html, September 2003.

[20] R. Rosson, "Converting between text strings and binary vectors," http://www. mathworks.com/matlabcentral/fileexchange/18072-converting-between-text-strings-and-binary-vectors, December 2007.

[21] J. Fridrich, "Structural LSB detectors," http://dde.binghamton.edu/download/structural_lsb_detectors/, November 2011.