DCT DOMAIN STEGASVM-SHIFTED LSB MODEL FOR HIGHLY IMPERCEPTIBLE AND ROBUST COVER-IMAGE

Saadiah Yahya[1], Hanizan Shaker Hussain[2], and Fakariah Hani M. Ali[1]

[1]*UiTM, Malaysia,*saadiah@tmsk.uitm.edu.my
[2]*Kolej Poly-Tech MARA, Bangi, Malaysia,* hanizan@gapps.kptm.edu.my
[1]*UiTM, Malaysia,* fakariah@tmsk.uitm.edu.my

ABSTRACT. The importance of information security in protecting and hiding information has increased due to the increased use of computers and Internet. Information hiding technology such as Digital Image steganography embeds secret messages inside other files. Least Square Bit (LSB) is the most popular technique used in image steganography that hides data behind a cover-image in a spatial and discrete cosine transform (DCT) domain. Support Vector Machine (SVM) is another technique that is used to strengthen the embedding algorithm. The main aim of image steganography is to keep the secret-message remain secret regardless of the techniques used. But many of the previously proposed techniques failed to attain this aim. The main concerns to this problem are the non-random changes of a cover-image that constantly occurred after the embedding process and the non-robustness of the embedding algorithm to image processing operation. This study therefore proposes a new model that utilises Human Visual System (HVS) and embedding technique through shifted LSB called SteganSVM-Shifted LSB in DCT domain to preserve the imperseptibility and increase the robustness of stego-images. The proposed technique shows better performances compared to other existing techniques.

Keywords: information security, secret messages, image steganography, least square bit, discrete cosine transform, support vector machine, human visual system

INTRODUCTION

Thousands of images files linger on the Internet every day and they may be exploited as a cover to hide secret-messages (Rabah, 2004; Davidson *et al.*, 2004). The importance of information security has increased due to the increased use of computers and the Internet which includes one of its exciting subfields i.e. image steganography (Cheddad*et al.*, 2010). Image steganography is the art and science of covert communication where the embedding of secret-message (e.g. text) in an innocuous-looking cover-image file (e.g. JPEG) is not known and sensible to human beings (Siponen & Oinas-Kukkonen, 2007).

BACKGROUND AND MOTIVATION OF THE RESEARCH

Many algorithms have failed to serve the main purposes of steganographic algorithm which is to keep the existence of the secret-message secret (Bin *et al.*, 2011; Cheddad *et al.* 2010). Secret-message is not robust enough to resist the passive steganalysis (e.g. image processing operation) (Cachin, 2004; Lee & Chen, 2000).

SVM has shown to be a good potential of overcoming most of the hiding problems (Hamel, 2009). Four rationales in choosing SVM are as follows. SVM is good in generalizing the non-linear problem (Hsu *et al.*, 2009); SVM can classify images and efficiently converges to a global optimum solution in a finite number of iterations (Ao*et al.*, 2010); The over-fitting problem can easily be controlled by the choice of support vectors and the kernel tricks (e.g. RBF or polynomial) (Parkes, 2011); and always seeks for optimal separating hyper-plane and margins that are maximally separated between the classes (Hamel, 2009; Kecman, 2005).

Problems Statements and Objectives

Existing image steganography techniques are facing some weaknesses. Firstly, is the non-random changes on a cover-image that constantly occurred after the embedding process (Cheddad*et al.*, 2010; Li *et al.*, 2006; Morkel*et al.*, 2005; Chandramouli*et al*., 2002). Secondly is the non-robustness of image steganography to image processing operation. Thus, it will bring distortion to the secret-message (Walia et al., 2010; Cheddad et al., 2010; Almohammad et al., 2008; Duric et al., 2005; Katzenbeisser and Petitcolas, 2000).

From literature, none of the existing image steganographic algorithms can address the problems stated**.** Therefore, this study proposed a conceptual SVM-Shifted LSB Image Steganographic Model based on DCT domain to preserve imperceptibility and robustness. The SVM-Shifted LSB Image Steganographic Algorithm will then be developed. Finally, the imperceptibility and robustness of the proposed algorithm will be evaluated and verified.

Scope of the Research

The study considers the following assumptions. The image domain used in this research is DCT domain. Digital gray scale image is considered as a cover-image: JPEG versus BMP. The text secret-message is used: Text versus Image. Only image steganography with statistical attacks i.e. passive steganalysis which takes action only when the stego-image is found suspicious considered. Imperceptibility versus Robustness versus Payload capacity will be tested via test-bed experiments.

Novelties and Significance of the Research

The research contributes four novelties. Firstly, the StegaSVM classification model which is based on SVM. It is the first and specifically designed for image steganography. Secondly, the StegaSVM-Shifted LSB Model that may highly preserves the imperceptibility and upholds the robustness. On the whole, three models are being introduced in this study: StegaSVM Classification; StegaSVM-Shifted LSB Embedding; and StegaSVM-Shifted LSB Extracting. Lastly, StegaSVM-Shifted LSB tool is developed to evaluate and validate the proposed algorithms.

The findings of this study can be leveraged by many in various areas. For example, in research and development field, the study may facilitate future research in image information hiding that emphasizes on imperceptibility and robustness. Further, it can strengthen the field of computer security and contribute to other related areas such as the medical field where patients' sensitive and confidential information can be hidden behind their image data (e.g. X-ray images (Petitcolas, 2000). The findings may also assist defense ministry in securing the national defense information. This can be done by embedding the information into image files such as digital location maps.

RELATED LSB AND SVM EFFORTS

LSB techniques proposed by Tseng & Chang 2004 and Nag *et al.* 2010 have been identified related to this study. Their techniques capitalise on imperceptibility, robustness and the use of HVS. However, the techniques are still vulnerable to the secret messages being easily detected. Other concerns are: secret-message is not compressed and encrypted; types of frequency to be used for embedding are not taken into account; HVS was not utilised as it should be; simply utilises each of the DCT coefficients from the whole cover-image; and more bits are embedded in one DCT coefficient.

SVM is a soft computing technique that has a better learning and generalization capabilities. This can be seen from the previous SVM efforts such as the Blind, SVM Watermarking which was propose by Meng et al., 2008, FSVM Watermarking which was done by Li et al., 2010, and SVM-SS Watermarking which was conducted by Ramly et al., 2011. However, these techniques are still vulnerable of secret messages being easily detected. Other concerns are: the non random embedding; only consider the non-smooth area; SVM is not implemented in the extracting process; the secret-message is not compressed and encrypted; and need the original cover-image to extract message.

In this study therefore, the statistical attack were deployed which take effect only when the stego-image is found suspicious.

RESEARCH METHODOLOGY

The methodological framework for this study is composed of five main phases which are done in sequence as shown in Figure 1.

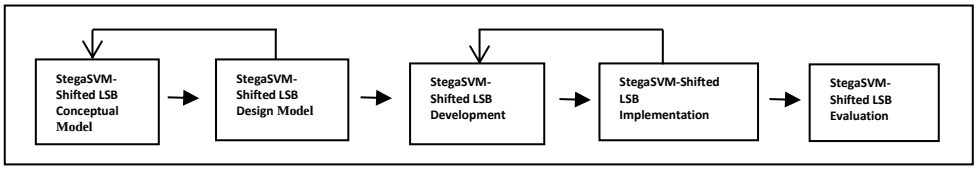

Figure 1. Steganographic Methodological framework

Each of the phases can be further deliberated with it key activities and deliverables. For example, the fifth phase, which is Stega-SVM Shifted Evaluation consists of five further activities shown in Figure 2.The phase is necessary to verify the reliability and validity of the model.

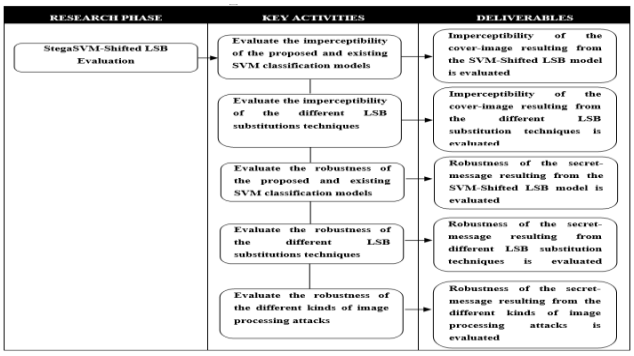

Figure 2. StegaSVM-Shifted LSB Evaluation Phase

RESULTS AND FINDINGS

Imperceptibility

Four types of cover-images of different image complexities have been used; Lena, Baboon, UiTM and Clock. As for the secret-message, the size that has been used is 1024bits for all types of cover-images. Mean while, for the cover-image classification, the value of *C* has been set to 30 while the value of γ is set at 0.5. Table 1 shows cover-images with high image complexities; Lena and Baboon have recorded the highest PSNR; 49.86 and 49.33 respectively. This is followed by UiTM cover-image with PSNR of 48.89 whereas the cover-image that has low image complexities, that is Clock, recorded the lowest PSNR with 47.68.

Hence, cover-image that has high image complexities is safer and suitable to be used in image classification - able to decrease distortion to the cover-image in the embedding process. The proposed model is successful in reducing the impact of non-random changes on a cover-image to avoid the secret-message from being easily detected.

Table 1. PSNR of Different Types of Cover-Images

Types of cover-	Different Image Com-	PSNR
Lena	High	49.86
Baboon	High	49.33
UiTM	Moderate	48.89
Clock	Low	47.68

Table 2 indicates, the increase in secret-message size leads to quality decline in certain cover-image. It is clearer that when the size of the secret-message is equal to 2048bits; the cover-image with the size of 256x256 could not afford to accommodate the secret-message.

Table 2. PSNR of various sizes of secret-messages

Types of cover-images	Different size of secret-message				
	128 bits	256 bits	512 bits	1024 bits	2048 bits
Lena	51.15	50.67	49.94	49.86	36.17
Baboon	50.09	49.81	49.76	49.33	33.65
UiTM	49.94	49.65	49.33	48.89	30.60
Clock	48.89	48.78	47.79	47.68	29.79
Average PSNR	50.22	49.72	49.20	48.94	32.55

Table 3. PSNR the proposed and existing models

Methods	Size of secret-message	Type of secret-message	PSNR
SVM- Shifted LSB	1024	Text	48.94
SVM-SS	1024	Text	41.97
Blind SVM	1024	Image	38.04

Table 3 shows the comparison of the suggested model against SVM-SS model and Blind-SVM. The comparison is made to secret-message that equals to 1024bits with secret-message of type text or image. StegaSVM-Shifted LSB model shows the highest PSNR of 48.94, while SVM-SS is the second followed by the Blind-SVM with the PSNR equals to 38.04.

This confirms that, StegaSVM-Shifted LSB model through StegaSVM classification technique, has taken the correct and appropriate image features into account (ie, luminance, edge and entropy) in exploiting HVS. This technique is also applied in the extracting process in order to extract the right secret-message and takes into account all areas of either smooth or non-smooth for the embedding process. The model not only has a better performance than other methods but has never been applied to any previous SVM techniques too.

Robustness

To determine the robustness, two experiments have been carried out. First, the robustness and second the multiple images processing attacks.

Table 4. NC of various sizes of secret-messages

	Various sizes of secret-messages (bits)				
	128	256	512	1024	2048
Normalized Cross-Correlation (NC)	1.00	1.00	1.00	1.00	0.87

In the first test, normalized cross-correlation (NC) value was determined from the experiment on different sizes of extracted secret-message. The size of that secret-message is from 128 until 2048bits. NC equals to the value of 1.0 for the secret-message size range from 128 to 1024bits. NC value decreased to 0.87 on extracted secret-message with the capacity of 2048bits. Table 4 illustrates the robustness of stego-image is better when the NC value of each secret-message is close to 1.0 for all cover-images. This confirms that the proposed model was able to maintain the authenticity and integrity of the extracted secret-message especially while in the extracting process.

Table 5. NC of the proposed and existing models

Methods	Normalized Cross-Correlation (NC)
SVM- Shifted LSB	1.00
SVM-SS	0.99
Blind SVM	1.00

Table 5 depicts the NC values which are affected after image processing attacks. It records that Blind SVM and SVM-Shifted LSB method attain a much higher value of NC compared to SVM-SS method.

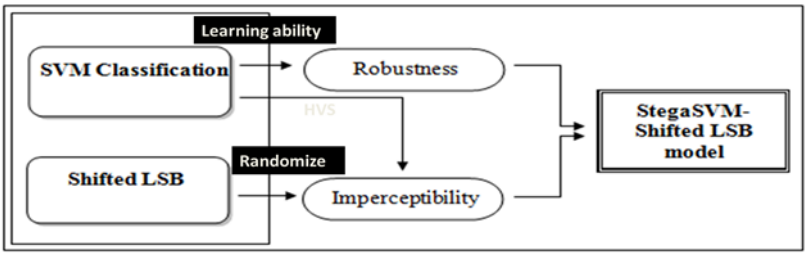

Figure 3. StegaSVM-Shifted LSB Model

Figure 3 summarises the proposed model which incorporates the following features: compress and encrypt secret-message before it is being used in the embedding process; utilise the image features (i.e. luminance, edge and entropy) in order to determine the potential area for embedding; utilise SVM technique in order to extract the right secret message; use JPEG

grayscale image that is known to be more robust medium; distribute secret-message randomly throughout the cover-image which make it more difficult to detect; and utilise the robust kernel, that is RBF.

CONCLUSION AND RECOMMENDATION

In conclusion, the proposed model using StegaSVM classification and Shifted LSB technique which utilises HVS and random embedding has truly increased the ability of a highly imperceptible and robust image steganographic model. For future works, researchers can utilise or explore a spatial domain. Further, the payload capacity as part of main requirement and the exploitation of the Discrete Wavelet Transform (DWT) can also be explored.

REFERENCES

Almohammad, A., Hierons, R. M., & Ghinea, G. (2008). High Capacity Steganographic Method Based Upon JPEG.*The Third International Conference on Availability, Reliability and Security*, 544-549. Barcelona: IEEE Computer Society Press.

Ao, S.-L., Rieger, B., & Amouzegar, M. A. (2010). *Machine Learning and Systems Engineering*. London: Springer.

Bin, L., Junhui, H., Jiwu, H., & Yun, Q. S. (2011). A Survey on Image Steganography and Steganalysis. *Journal of Information Hiding and Multimedia Signal Processing*, 142-172.

Cachin, C. (2004). An Information-Theoretic Model for Steganography. *Information and Computation*, 41-56.

Chandramouli, R., Grace, L., & Nasir, M. (2002). Adaptive Steganography. *Conference on Security and Watermarking of Multimedia Contents*, 69-78. San Jose: SPIE.

Cheddad, A., Condell, J., Curran, K., &Kevitt, P. M. (2010). Review: Digital Image Steganography: Survey and Analysis of Current Methods. *Journal Signal Processing*, 727-752.

Davidson, I., Paul, G., & Ravi, S. S. (2004). Steganography Using Spatially Interesting Pixels. *Lecture Notes in Computer Science*.

Duric, Z., Richards, D., & Kim, Y. (2005). Minimizing the Statistical Impact of LSB Steganography. *International Conference on Image Analysis and Recognition*, 1175–1183. Toronto: Springer.

Hamel, L. (2009). *Knowledge Discovery with Support Vector Machines*. New Jersey: John Wiley & Sons Inc.

Hsu, C. W., Chang, C. C., & Lin, C. J. (2009). A Practical Guide to Support Vector Classification. *Bioinformatics*, 1-15.

Katzenbeisser, S., & Petitcolas, F. A. (2000). *Information Hiding Techniques for Steganography and Digital Watermarking*. Boston: Artech House Inc.

Kecman, V. (2005). Basics of Machine Learning by Support Vector Machines .*In V. Kecman, Studies in Fuzziness and Soft Computing* (pp. 49-103).Cornell University.

Lee, Y. K., & Chen, L. H. (2000). A Secure Robust Image Steganographic Model.*The Conference on Information Security*, 275-284. Hualien, Taiwan.

Li, L., Ding, W.-Y., & Li, J.-Y. (2010). A Novel Robustness Image Watermarking Scheme Based on Fuzzy Support Vector Machine. *The 3rd IEEE International Conference on Computer Science and Information Technology*, 533 – 537. Chengdu: IEEE Xplore.

Li, Q., Yu, C., & Chu, D. (2006). A Robust Image Hiding Method Based on Sign Embedding and Fuzzy Classification. *The 6th World Congress on Intelligent Control and Automation*, 10050-10053. Dalian, China: IEEE Computer Society.

Meng, F., Peng, H., Pei, Z., & Wang, J. (2008). A Novel Blind Image Watermarking Scheme Based on Support Vector Machine in DCT Domain. *International Conference on Computational Intelligence and Security*, 16 - 20. Suzhou: IEEE Xplore Digital Library.

Morkel, T., Eloff, J. H., & Olivier, M. S. (2005). An Overview of Image Steganography. *Annual Information Security South Africa Conference*.Sandton, South Africa: M S Olivier.

Nag, A., Biswas, S., Sarkar, D., & Sarkar, P. P. (2010). A Novel Technique for Image Steganography Based on Block-DCT and Huffman Encoding. *International Journal of Computer Science and Information Technology*, 103-112.

Parkes, D. C. (2011, January). Overfitting. *Lecture Notes*, pp. 53-96. Harvard University.

Petitcolas, F. (2000). *Introduction to Information Hiding*. Norwood: Artech House.

Rabah, K. (2004). Steganography-The Art of Hiding Data. *Information Technology Journal*, 245-269.

Ramly, S., Aljunid, S. A., & Hussain, H. S. (2011). SVM-SS Watermarking Model for Medical Images. *International Journal of Digital Enterprise and Information Systems*, 372-386.

Siponen, M. T., & Oinas-Kukkonen, H. (2007). A Review of Information Security Issues and Respective Research Contributions. *Database for Advances in Information Systems*, 60-80.

Tseng, H.-W., & Chang, C.-C.(2004). High Capacity Data Hiding in JPEG-Compressed Images. *International Journal of Informatica*, 127-142.

Walia, E., Jain, P., & Navdeep. (2010). An Analysis of LSB and DCT Based Steganography. *Global Journal of Computer Science and Technology*, 4-8.