



LARGE DATA EMBEDDING; PROBLEM & SOLUTION

Akubuilu, K¹, Rix Torimiro²

^{1,2}Faculty of Engineering, University of Zimbabwe, Harare, Zimbabwe

*Corresponding Author: Akubuilu K

Article Received: 28-06-19

Accepted: 06-09-19

Published: 26-12-19

Licensing Details: Author retains the right of this article. The article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licences/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the Journal open access page

ABSTRACT

The objective of the study was to propose a high capacity data embedding system from DCT domain. The default quantization table is used for fixed mask projection quantization steps. Accordingly, DCT coefficient of the host image from low to higher frequency bands and embed bits from band to band with specially designed base vectors called Hadamard vectors. The embedding scheme procedure is described step wise. The proposed system can be used for data embedding and can be associated with another encryption method to make it safer. The proposed system is reported to be highly robust and secure.

Keywords: Embedding, Data, Large, DCT, Image.

INTRODUCTION

Digital form is frequently used in the 21st century for transferring data such as audio, video, pictures, and text files. A related concept is the data hiding which is about hiding data in some cover medium in order to avoid the misuse of data. Similarly, data embedding is used for digital contents authentication, secret data hiding, and copyright control. Based on their functionalities, data embedding techniques are classified as steganographic or watermarking. The steganographic technique is considered as a superior technique since data can be hide in different formats which is statistically and perceptually undetectable. On the other hand, the embedding data method is about replacing bits of unused or useless data in the source cover for data to be undetectable. Spatial embeddedness and transform embeddedness are two common types of the embeddedness. The spatial embedding is that type where messages are inserted into the LSBs of image pixels. Whereas, the transform embedding is about a message embedded by modifying

frequency coefficients of the cover image. Generally, the robust method is considered as superior to the spatial embedding methods since it is considered as safer and less prone to the various types of attacks. However, with steganography technique, the perceptibility, is a critical quality. The capacity of steganography is also an issue. The capacity issue is about how much information can be embedded relative to its perceptibility. In this study, the cover object will be used as digital image in which information is hidden. The challenge related to the embedding method for data in cover images is to hide as much data as possible with the least noticeable difference in the output image and to obtain higher robustness.

The aim of the paper is to propose a high capacity data embedding system which belongs to the DCT domain. In this study, the DCT is about Hadamard matrix and the Project Quantization technique for data embedding.

The background of the high capacity data embedding is that there is a lot of variety in the domain of high images data processing (Bender, Gruhl, Morimoto, & Lu, 1996; Petitcolas, Anderson, & Kuhn, 1999). A small data is encoded for maintenance of secrecy of data (Kaur & Ranade, 2012). We propose the least-significant-bit method for obtaining higher payloads (Wang, Lin, & Lin, 2001; Chan & Cheng, 2004). These methods typically utilize some mapping rules to embed the important image in certain LSB plans for covering image and applying additional pixel-adjustment procedure for reducing the error introduced in the embedding process. Some authors such as Lee & Chen (2000) suggest that one should consider the human vision system characteristics for evaluation of the number of bits which may be hidden in an image. Based on this, more information is hidden in spatial variations as human eyes have higher sensitivity to edges. For steganography system, two common properties are imperceptibility and the data payload (Zhang & Wang, 2004). These two requirements have opposing nature as higher payload brings reduction in the perceptibility of the hidden data. Therefore, there is required to be a trade-off between the visual imperceptibility of the data embedding system and the data capacity. Imperceptibility metric is about an attempt to balance the both requirements. Mostly in commercial settings, there is requirement of large amount of communication for exposure of its value and the position. However, because of security reasons, in some settings, there is low bandwidth available to a communication channel. The paper presents a least-significant-bit replacement and pixel-value method for providing an imperceptible stego-image quality. If PVD method is used in isolation, it can store larger information with good visual quality of stego-image. Other possible methods are also proposed by different experts such as Pseudo Random Number Generator, Discrete Cosine transform, and Vector Quantization. One steganography method is the utilization of two block-matching (Wang & Chen, 2006). The method uses highest similarity block for each block of the important image. The indexes and bases are not-well-matched blocks are recorded in the least significant bits of the cover image using a hop scheme. This method shows higher transmission time requirements, reduced storage and higher data payload. Based on such limitations, there are many scholars who proposed hiding capacity of information hiding algorithm based on DCT transformation in order to extend its application. Lan & Tewfik (2006) developed an efficient high-capacity embedding algorithm which utilizes

the quantized projection embedding method. The system is good since it provides suitable balance between visual quality, data capacity and robustness. The hiding information in low frequency of DCT is found to be more robust compare to the higher frequency due to the sensitivity of signals by human vision system. JPEG which commends a standard quantization matrix, higher frequency leads to the higher quantization value thus information embedded in higher frequency will be easily filtered by JPEG compression. Accordingly, information hiding algorithm can be used for embedding information in DCT median and higher frequency coefficients which shows higher robustness against heavy compression (Xie, Xie, & Huang, 2009). Based on LSB substitution method, four-pixel differentiation method is implemented.

The Projection Quantization Technique

In this quantized projection, $Q_0(p) = \Delta Q_0(BH^T v / \Delta)$, is equal to $p + \epsilon p$. The random variable ϵp is the quantized projection error term. The distribution function of the error term can be modeled as a uniform function within the range $L - \Delta, \Delta$ where L and mean exclusive and Δ is the quantization step used. Superscript H indicates the hermitian (or transpose for a real vector) and B denotes the base vector used for projection, with

$$B = [b_1, b_2, \dots, b_N]^T.$$

Vector B is a normalized vector, so its norm is 1. Accordingly, BB^H refers to the projection matrix of the subspace of B while, Q_0 and Q_1 refers to the closest odd and even numbers it can be noted that most vectors are $N \times 1$ dimensions. If we substitute, $p + \epsilon p$ in equation 1, we can obtain expressed in terms of ϵp .

$$\tilde{v} = (I - P)v + B(p + \epsilon p) \quad (2)$$

$$\tilde{v} = v + \epsilon p B. \quad (3)$$

In situation of quantization attacks on embedded block vector, the resulting quantized will be \tilde{v} , $Q_c(\tilde{v})$ as

$$Q_c(\tilde{v}) = v + \epsilon p B + E_q \quad (4)$$

In the above equation, E_q is the quantization error random vector with associated quantization steps δ is.

The Proposed Method

The default quantization table is used for fixed mask projection quantization steps. Accordingly, DCT coefficient of the host image from low to higher frequency bands and embed bits from band to band with specially designed base vectors called Hadamard vectors. The table is as follows.

Table 1: Proposed Matrix

26	21	20	26	24	40	61	71
22	22	24	29	36	69	71	65
16	17	17	34	46	70	81	72
16	32	32	39	47	79	103	79
19	37	39	66	49	90	109	98
26	59	65	74	53	107	121	102
51	91	81	97	67	123	134	105
83	97	99	99	87	131	138	109

The embedding scheme procedure is described below;

First, we have to calculate the coefficients of 8 by 8 blocks of the host image. Then the coefficients need to be rearranged from lowest frequency to the highest one based on its band. Frequency band is having their own corresponding Qtable. Next Hadamard matrix is used for allowing multiple bit embedding. The frequency of the bands needs to be sorted by utilizing the Qtable from the smaller to larger value with the exception of DC band which is always remains the first band. For performing the multi-bit embedding unit base vectors, embed 0 or 1 for even and odd quantization operators to each block vector of the sorted frequency bands. For instance, for an image of 256 by 256, the image frequency band size will be 32 by 32. The column vectors of a frequency band are the block vector. Even number quantization is used for embedding bit '0'.

$$\tilde{v} = v - Pv + B\Delta Q0(BHv/\Delta).$$

For the odd number quantization is used for embedding bit '1'

$$\tilde{v} = v - Pv + B\Delta Q1(BHv/\Delta).$$

For reconstruction of the image, once all embedding is done, reverse reordering of the DCT coefficients is performed based on embedding process from band to band. Inverse discrete cosine transform is applied for modification and correction of ordered DCT coefficients for producing the embedded image. At any range, the embedding process can be stopped based on user's preference for visual distortion. Addition of more bits means more distortion is created.

For extraction process, it can be said that it is similar to the last part of the embedding process with the exception that we need to calculate the projection (BH \tilde{v}) instead of embedding bits. Based on corresponding Δ , computed projection is rounded up. The extracted bit can be 0 or 1 based on the result. Additionally, encryption method such as md5 or RSA can be used to enhance the security of the data.

CONCLUSION

Embedding data in spatial domain have better desired features in comparison to other techniques such as higher security and the larger capacity. In this study, we proposed and tested a system which can be used for embedding larger data in DCT domain with minimum distortion. It can be argued that this proposed system is highly secure and robust thus making it a desirable system of choice. The system also shows less BER under JPEG attacks.

References

- Bender, W., Gruhl, D., Morimoto, N., & Lu, A. (1996). Techniques for data hiding. *IBM systems journal*, 35(3.4), 313-336.
- Chan, C. K., & Cheng, L. M. (2004). Hiding data in images by simple LSB substitution. *Pattern recognition*, 37(3), 469-474.
- Kaur, N., & Ranade, S. K. (2012). High Capacity Data Embedding System in DCT domain for Colored Images. *International Journal of Computing and Business Research (IJCBR)*, 3(3).

- Lan, T. H., & Tewfik, A. H. (2006). A novel high-capacity data-embedding system. *IEEE Transactions on Image Processing*, 15(8), 2431-2440.
- Lee, Y. K., & Chen, L. H. (2000). High capacity image steganographic model. *IEE Proceedings-Vision, Image and Signal Processing*, 147(3), 288-294.
- Petitcolas, F. A., Anderson, R. J., & Kuhn, M. G. (1999). Information hiding-a survey. *Proceedings of the IEEE*, 87(7), 1062-1078.
- Wang, R. Z., & Chen, Y. S. (2006). High-payload image steganography using two-way block matching. *IEEE Signal Processing Letters*, 13(3), 161-164.
- Wang, R. Z., Lin, C. F., & Lin, J. C. (2001). Image hiding by optimal LSB substitution and genetic algorithm. *Pattern recognition*, 34(3), 671-683.
- Xie, J., Xie, Q., & Huang, D. (2009). A robust high capacity information hiding algorithm based on DCT high frequency domain. In *2009 International Symposium on Computer Network and Multimedia Technology* (pp. 1-4). IEEE.
- Zhang, X., & Wang, S. (2004). Steganography using multiple-base notational system and human vision sensitivity. *IEEE signal processing letters*, 12(1), 67-70.