Oriental Journal of Computer Science & Technology Vol. **3**(1), 01-04 (2010)

JPEG compression steganography

AKHIL KHARE[1], MEENU KUMARI[1] and PALLAVI KHARE[2]

[1]Department of IT,Bharati Vidyapeeth, University & College of Engineering, Pune - 411 043 (India).
[2]Department of E & TC, SSSIST, Bhopal - 462 039 (India).

(Received: April 12, 2010; Accepted: June 04, 2010)

ABSTRACT

In any communication, security is the most important issue in today's world. Lots of data security and data hiding algorithms have been developed in the last decade, which worked as motivation for our research. In this project, named "JPEG Compression Steganography", we have designed a system that will allow an average user to securely transfer text messages by hiding them in a digital image file using the local characteristics within an image. This project is a combination of steganography and encryption algorithms, which provides a strong backbone for its security. The proposed system not only hides large volume of data within an image, but also limits the perceivable distortion that might occur in an image while processing it. This software has an advantage over other information security software because the hidden text is in the form of images, which are not obvious text information carriers. The project contains several challenges that make it interesting to develop. The central task is to research available steganography and encryption algorithms to pick the one the offer the best combination of strong encryption, usability and performance. The main advantage of this project is a simple, powerful and user-friendly GUI that plays a very large role in the success of the application.

Keywords: JPEG Compression Steganography.

INTRODUCTION

The past decade has witnessed a surge of research activity in multimedia information hiding, targeting applications such as steganography[1], digital rights management, and document authentication. Numerous works are available in the literature related with the Image Steganography. We propose a framework for hiding large volumes of data in images while incurring minimal perceptual degradation. The embedded data can be recovered successfully, without any errors, after operations such as decompression, additive noise, and image tampering. The proposed methods can be employed for applications that require high-volume embedding with robustness against certain non-malicious attacks. The hiding methods we propose are guided by the growing literature on the information theory of data hiding (summarized in the next paragraph), but are adapted to the specific application of hiding in images.

Information-theoretic treatments of the data hiding problem typically focus on hiding in independent and identically distributed Gaussian host samples. The hider is allowed to induce a mean squared error of at most $D1$, while an attacker operating on the host with the hidden data is allowed to induce a mean squared error of at most $D2$. Information-theoretic prescriptions in this context translate, roughly speaking, to hiding data by means of the choice of the vector quantizer for the host data, with the AWGN attack being the worst-case under certain assumptions. This method of hiding was first considered by Costa , based on results of Gel'fand and Pinsker on coding with side information. Game-theoretic analyses of data hiding, with the hider and attacker as adversaries, have been provided by Moulin and O'Sullivan, and by Cohen and Lapidoth. Estimates of the hiding capacity of an image, based on a parallel Gaussian model in the transform domain, have been provided by Moulin and Mihcak. Chen and Wornell present a variety of practical approaches to data hiding, with

a focus on scalar quantization based hiding, and show that these schemes are superior to spread spectrum hiding schemes, which simply add a spread version of the hidden data to the host. A scalar quantization based data hiding scheme, together with turbo coding to protect the hidden data, while a trellis coded vector quantization scheme is considered by Chou et al.

Though the concept of steganography and cryptography are the same, but still steganography differs from cryptography. *Cryptography*[2] focuses on keeping the contents of a message secret, *steganography* focuses on keeping the existence of a message secret. Steganography and cryptography are both ways to protect information from unwanted parties but neither technology alone is perfect and can be compromised. Once the presence of hidden information is revealed or even suspected, the purpose of steganography is partly defeated. The strength of steganography can thus be amplified by combining it with cryptography.

Methodology

Relative to the preceding methods, a key novelty of our approach is that our coding framework permits the use of local criteria to decide where to embed data. The main ingredients of our embedding methodology are as follows.

(a) As is well accepted, data embedding[3] is done in the transform domain, with a set of transform coefficients in the low and mid frequency bands selected as possible candidates for embedding.

(b) A novel feature of our method is that, from the candidate set of transform coefficients, the encoder employs local criteria to select which subset of coefficients it will actually embed data in. The use of local criteria for deciding where to embed is found to be crucial to maintaining image quality under high volume embedding.

(c) For each of the selected coefficients, the data to be embedded indexes the choice of a scalar quantizer for that coefficient. We motivate this by an information theoretic analysis showing that, for an idealized model , scalar quantization[2] based hiding is only about 2 dB away (in terms of resilience to attack) from optimal vector quantization based hiding.

(d) The decoder does not have explicit knowledge of the locations where data is hidden, but employs the same criteria as the encoder to guess these locations. The distortion due to attacks may now lead to insertion errors (the decoder guessing that a coefficient has embedded data, when it actually does not) and deletion errors (the decoder guessing that a coefficient does not have embedded data, when it actually does). In principle, this can lead to desynchronization of the encoder and decoder.

(e) An elegant solution based on erasures and errors correcting codes is provided to the synchronization problem caused by the use of local criteria. Specifically, we use a code on the hidden data that spans the entire set of candidate embedding coefficients, and that can correct both errors and erasures. The subset of these coefficients in which the encoder does not embed can be treated as *erasures at the encoder.* Insertions now become errors, and deletions become erasures (in addition to the erasures already guessed correctly by the decoder, using the same local criteria as the encoder). While the primary purpose of the code is to solve the synchronization problem, it also provides robustness to errors due to attacks.

Two methods for applying local criteria are considered. The first is the block-level *Entropy Thresholding (ET)* method, which decides whether or not to embed data in each block (typically 8X8) of transform coefficients, depending on the entropy, or energy, within that block. The second is the *Selectively Embedding in Coefficients (SEC)* method, which decides whether or not to embed data based on the magnitude of the coefficient. *Reed-Solomon (RS)*[2] codes are a natural choice for the block-based ET scheme, while a “turbo-like” *Repeat Accumulate (RA)*[2] code is employed for the SEC scheme. We are able to hide high volumes of data under both JPEG and AWGN[2] attacks. Moreover, the hidden data also survives wavelet compression, image resizing and image tampering attacks. The block diagram of the proposed system is shown in Figure 1.

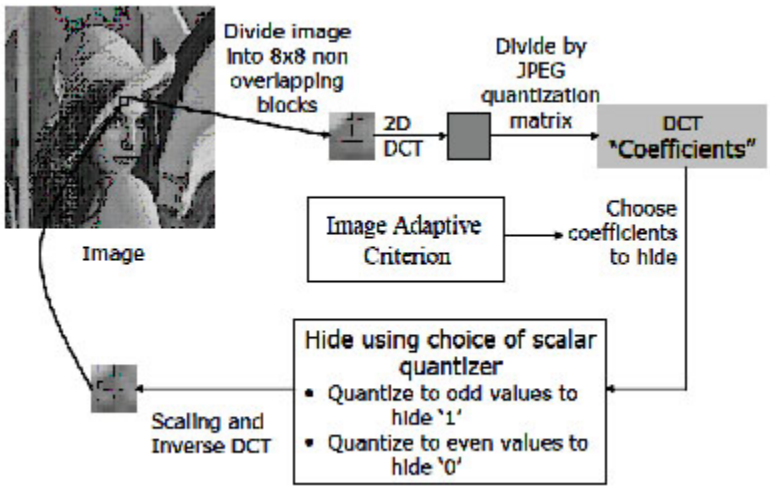
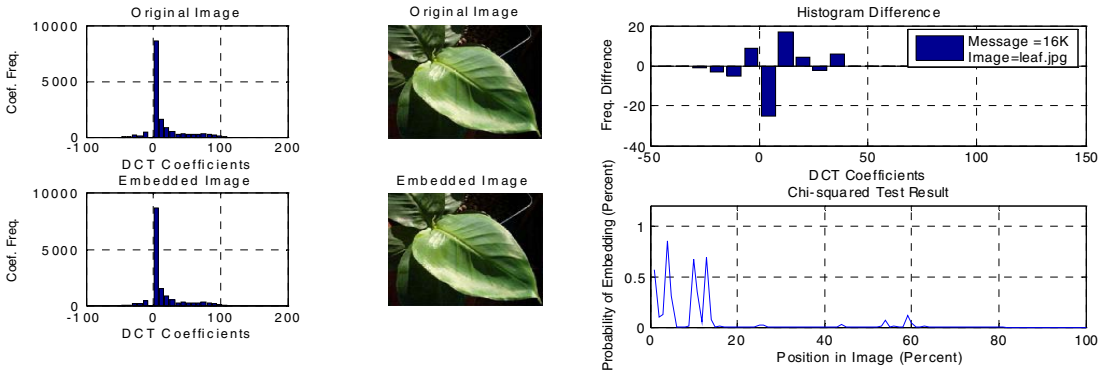


Fig. 1: Image-adaptive embedding methodology

RESULTS

All steganographic algorithms have to comply with a few basic requirements. The requirements are: Invisibility, Payload capacity, Robustness against statistical attacks, Robustness against image manipulation, Independent of file format and Unsuspicious files. Unfortunately, there is not one algorithm that satisfies all of the requirements. Thus a trade-off will exist in most cases, depending on which requirements are more important for the specific application.

The process of embedding information during JPEG compression results in a stego image with a high level of invisibility, since the embedding takes place in the transform domain. JPEG is the most popular image file format on the Internet and the image sizes are small because of the compression, thus making it the least suspicious algorithm to use. However, the process of the compression is a very mathematical process, making it more difficult to implement. The JPEG file format can be used for most applications of steganography, but is especially suitable for images that have to be communicated over an open systems environment like the Internet. The result analysis of the proposed system is shown in Fig. 2.



Embedding Message of 16 Kbytes leaf.jpg

Fig. 2: Histogram Difference & Chi-Squared Result of Original & Embedded Image

CONCLUSION

The meaning of Steganography is hiding information and the related technologies. Steganography and Encryption can be applied together for better security. This system allows a user to securely transfer a text message by hiding it in a digital image file. 128 bit AES encryption is used to protect the content of the text message even if its presence were to be detected. Currently, no methods are known for breaking this kind of encryption within reasonable period of time. Additionally, compression is used to maximize the space available in an image.

ACKNOWLEDGEMENTS

The work on this paper was supported by the Bharati Vidyapeeth University & College of Engineering, Pune. The views and conclusions contained herein are those of the authors and the paper contains the original work of the authors. We took help from many books, papers and other materials.

REFERENCES

1. N . Provos and P. Honeyman, "Hide and Seek: An introduction to Steganography," *IEEE Security & Privacy Journal.* (2003).
2. Ranjan Bose, "Information Theory Coding and Cryptography".
3. Steven W. Smith, The Scientist and Engineer's Guide to Digital Signal Processing
4. H.Ancin, Anoop K.Bhattacharjya, Joseph Shu, "Improving void-and-cluster for better halftone uniformity",International Conference on Digital Printing Technoogies.
5. Katzenbeisser and Petitcolas , "Information Hiding Techniques for Stenography and Digital watermaking" Artech House, Norwood, MA. 2000 .
6. C.Chang, C.Tsai, and T.Chen, "A new scheme for sharing secret color images in computer network", in Proc. of International Conference on Parallel and Distributed Systems, 2000, pp. 21-27.
7. R.L.Alder, B.P.Kitchens, M.Martens, "The mathematics of halftoning", *IBM J. Res. & Dev.,* **47(**1), pp. 5-15 (2003).
8. R.Lukac, K.N.Plantaniotis, B.Smolka, "A new approach to color image secret sharing", *EUSIPCO,* pp.1493-1496 (2004).
9. Fridrich , J ., Goljan M., and Hogea , D ; New Methodology for Breaking stenographic Techniques for JPEGs. " Electronic Imaging (2003).