Robust Source Coding Steganographic Technique Using Wavelet Transforms

S. K. Muttoo[1] and **Sushil Kumar**[2]

Abstract - Information hiding has emerged as an important research field to resolve the problems in network security, quality of service control and secure communications through public and private channels. Keeping the network in a desired state is the utmost requirement of network communications. The work is being done in different fields to achieve this goal. Strganography is one of the branches of information hiding that is used to solve this problem. In this paper we present a Steganographic algorithm based on wavelet transforms. Our algorithm first uses the Best T-codes to encode the message before embedding into a cover image. The one of the advantage of this is that we can embed high capacity messages into the cover objects. The second advantage of using T-codes is self-synchronization attained at decoding stage. To achieve better imperceptibility of stego-image, we have embedded the encoded message into the cover image using wavelet fusion technique more than once, by selecting each time the wavelet block pixels using the pseudo random permutations. From the experimental results we have observed that the algorithm is imperceptible and can have 100% embedding capacity.

Index Terms - Steganography, SSVLC, DWT, PSNR

1. INTRODUCTION

In this information era, either a public network or private network, one requires a tool that can allow communicating over these channels and as well providing the security and robustness of the hiding data. The information hiding has emerged as a useful and important field for resolving the problems of public network security and secure communications. There are three main streams of research areas over which this field is focused at present and they are *Steganography, Watermarking and Cryptography*. In Cryptography, the data is encrypted so that it cannot be understood by anyone else. The encrypted data is unreadable but is not hidden from the eavesdroppers. Though the purpose of *Cryptography* is to protect the data (or information) from unwanted attackers, it does not ensure covertness on the channel. The Steganography solves this problem by embedding data in the cover object so that it is hard to detect. The branch of Watermarking is to embed a watermark for the purpose of copyright protectction, authentication and temper proofing.

There are mainly four requirements of any information hiding technique, namely, Imperceptibility, Capacity, Security and Robustness. Imperceptibility means that human eyes cannot distinguish the difference between the steg-image and the original image. Capacity refers to the amount of data that can be embedded in the cover object. Security means that an eavesdropper cannot detect the hidden data, and Robustness requires that the hidden data can be recovered within certain acceptable errors even when the steg-image has endured some signal processing or noises.

Now-a-days Cryptography or Source encoding methods have also been used in conjunction with *Steganography* to provide an additional layer of security. Over time the information hiding techniques have improved to meet the desired goal. Digital steganography provides privacy for intelligence and military personnel and for people who are subject to censorship.

There are various domains of information hiding viz., *spatial domain*, t*ransform domain and spread spectrum domain*. The transform domain based hiding techniques has not only the potential to achieve higher capacity than the spatial domain based techniques, they are also found to be more robust.

Apart from text, images have been used widely as cover objects for the purpose of information hiding as their digital representation provide high degree of redundancy. The most popular transform hiding techniques Steganography systems are based on *discrete Fourier transform* (DFT), *discrete cosine transform* (DCT), *discrete wavelet transform* (DWT), *singular value decomposition* (SVD) *transform* and *discrete Hadamard transform* (DHT). These techniques are independent of an image formats and hide data in more significant areas of the transformed image. The details about these techniques can be found in [1-3, 9,10, 19, 21, 29].

In this paper we present a Steganographic method based on wavelet transform. We have first used best self-synchronizing T-codes to encode the original text. The purpose of using the T-codes is lying in the inherent self-synchronizing property of T-codes. According to [25], T-codes require anything between 1.5 to 3 symbols to attain synchronization following a lock loss. Also, by sending the message in the cover image in compressed form increases its security as well as embedding capacity. The secret message is then embedded into the cover image using wavelet-fusion technique [26] with a stego-key. To increase the quality (hence, PSNR value) of the stego-image to meet the imperceptible attribute of the steganography we embed the message in the cover image number of times but each time we use pseudo random number generator to select the pixel locations in the block. In the extracting algorithm we obtain the hidden message by taking the average of the messages extracted from the stego-image using the stego-key. To

[1]*Reader, Department of Computer Science, University of Delhi, Delhi, India*
[2]*Reader, Rajdhani College, University of Delhi, New Delhi ,India*
E-Mail: [1]*skmuttoo@cs.du.ac.in and*
[2]*azadsk2000@yahoo.co.in*

check the robustness of the algorithm, we have analyzed our algorithm against noise such as Salt and Pepper, Gaussian and Speckle and found satisfactory results.

2. SELF-SYNCHRONIZING VARIABLE LENGTH CODES

The categories of coding that minimize redundancy of information are Entropy coding, Source coding and Hybrid coding. Entropy coding is a lossless process whereas source coding is a lossy process. Most multimedia systems apply Hybrid coding techniques. The popular *variable length* most *codes* (VLC) for loss less compression used is Huffman codes. However, when an uncorrected error occurs in the encoded data it may propagate to the extent that all subsequent data are lost. Thus, one requires VLC with the property that data may resynchronize automatically after an error occurs in a minimum delay. There can be another problem of slippage which occurs However, if the number of symbols decoded before resynchronization are found to be different from the actual number of data symbols which have been encoded, raises the problem known as *Slippage problem* [16]. The slippage problem may leads to misinterpretation of the remaining data that howsoever may have been received correctly.

There are number of methods proposed to find the solution of synchronization problem. Some of the proposed techniques used restart markers but they increases overhead, i.e., bit rate. Thus, researchers realized that the VLC that provides the synchronization without the increase in overhead is needed. Gavin R. Higgie [7], Mark R. Titchner [23] and A.C.M. Fong [5] proposed a self-synchronizing VLC, viz., T-code. According to Titchener [25] , T-codes re-synchronize within one to three code words. G. Ulrich [27] and P.Reddy [20] have shown that T-codes exhibit better synchronizatjion properties when compared to Huffman codes. A.C.M. Fong et al have proposed the application of minimal sync-delay T-codes for information source coding. G.Y. Hong et al [8] have also investigated the application of self-synchronizing VLC (SSVLC).

S.K.Muttoo and Sushil kumar [11-13] have shown the application of Best T- codes in the two popular steganographic algorithms, Jpeg-Jsteg [28] and OutGuess 0.1 [17].

A. T-codes

T-codes are families of VLCs that exhibit extraordinarily strong tendency towards self-synchronization. The concept of 'simple T-codes' was given by M.R.Titchner[23]. He proposed a novel recursive construction of T-codes known as the '*Generalized T- codes'* that retain the property of self-synchronization [24]. Each T-augmentation step is characterized by two parameters: *a 'T-prefix' p*, a codeword from the existing T-code and *a 'T-expansion parameter' k,* a positive integer. Starting at augmentation level 0 with initial set S= {0, 1}, the construction of T-codes at augmentation level 1, 2 and 3 are summarized in the table below:

There can be many possible code sets matching a source depending on the parameters (p, k) chosen [24]. Apart from the generalized class of self-synchronizing efficient codes, T-codes show the best synchronization performance amongst the most efficient VLC's and require anything between 1.5 to 3 characters to attain synchronization following a lock loss. Among the subgroups of T-codes, the search for a best T-code set means those T-code sets that are optimally efficient and at the same time exhibits the least synchronization delay. Different T-codes exhibit different degree of synchronization performance, even if they have the same average code word length. The *Expected (or Average) synchronization delay* (ESD or ASD) is normally used as measure of synchronization performance. The ESD is defined as the average number of symbols in S that the decoder has to receive before it can conclude that it has achieved synchronization with respect to its largest level set. A number of attempts have been made to quantify the synchronization performance of different T-codes [25, 27, 5].

Ulerich Gunther [27] in his thesis has given a recursive search algorithm that yields the T-codes set with the minimum redundancy for a given source. This search algorithm utilizes equivalence and feasibility criteria to significantly restrict the search space. The best T-codes used in our algorithms in this paper are based on the breadth-first search algorithm proposed by Ulrich Gunther [27]. Ulrich chooses the least redundant set from a pool of all possible T-code sets by calculating redundancy for each of them. The search process is optimized by certain proposed constraints. The algorithm returns a group of code sets with least redundancy. To choose the best code set with least synchronization delay, we test each code set against very long test message string (composed of source symbols) by calculating ESD.

3. THE PROPOSED STEGANOGRAPHIC ALGORITHM

A large number of image Steganographic methods have been proposed over the last few years to achieve better perceptibility, best data hiding rate, survivability and security. The most of these embedding algorithms in a transform domain make use of DFT, DCT, DWT or DHT. Eric A. Silva and Sos S. Agaian [22] have embedded data in different transform domains and observed that the Haar wavelet transform is the best choice as compare to FFT, DCT or DHT for their method. However they observed that the relative performance of each of the transforms used were uniform across all images tested.

Our proposed method is a high capacity image steganogrphic method using Wavelet-fusion- method proposed by M. Fahmy Tolba and Al-said Ghonemy [26]. The proposed algorithm consists of four parts: Encoding, Embedding, Extraction and Decoding. Our algorithm

provides multi-level securities. First in encoding stage, we apply Best T-codes on the message for source coding. An encoded key is used for this purpose. The secret (encoded) message is then embedded in the cover image using wavelet-fusion-technique. To enhance the quality of stego-image we have embedded the message in the cover image number of times. The stego-key is used to select random pixels for embedding message. We require the stego-key to extract the hidden message. Finally, in the decoding stage, the original message is obtained with the help of encoded key. The steps of these algorithms are described in the figures 3.1 and 3.2.

The Embedding algorithm can be summarized as follows:

0. Input the Cover image and original text (or message)
1. Normalize the cover image. i.e., the pixel values made to lie between 0.0 and 1.0.
2. Apply preprocessing on cover image: choose 'alpha' (preferably between 0 and 0.1) and reconstruct pixels to lie in the range [alpha, 1 – alpha]. This will ensure that pixels from the fused coefficients (during embedding) would not go out of range and hence the secret message will be recovered correctly.
3. Apply 2D Haar transform on each color plane separately.
4. Encode the original message using best T-codes. The resulting secret message is a bit-stream of 0 and 1, denoted by (m_1 m_2…. m_n), where n is the embedding message length.
5. Generate pseudorandom permutation, using a stego-key, of the size equal to the length of cover image.
6. Enter the number of times the message to be embedded, num.
7. for i = 1 to num do

7.1 Select wavelet coefficient of the transformed image randomly, say f(j, k)

7.2 Embed the secret message bit, m (i), into the transformed image in the following way:

```
if   m(i) = '1'
   f (j,k) = f (j,k) + alpha;
else
   f (j,k) = f (j,k) – alpha;
```

8. Apply the inverse 2D Haar transform on each color plane separately.
9. Denormalize the image
10. Output: the Stego-image.

The Extraction algorithm is just the reverse process of the embedding method. We can summarize it as follows:

1. Apply 2D Haar transform on each color plane of the stego-image
2. Enter num, number of times message bwing embedded
3. Initialize the hiddenmessage to zero.
4. for j= 1 to num do

4.1 Select the embedded coefficients, i, using the PRNG based on the stego-key same as used in the embedding procedure.

4.2 Extract the embedded value of alpha by subtracting the original cover image from the stego image in the wavelet domain.

4.3 Obtain the secret message bit, m(i) as follows:

```
If alpha >0
   m(i) = '1'
else
   m(i)='0';
```

5. hiddenmessage += m(i);
6. end; /*for(j)*/
7. hiddenmessage /=num;
8. Decode the hiddenmessage using best T-code using the encoded key.

Output: Original message.

4. EXPERIMENTAL RESULTS

For testing our algorithm we have used 256 x 256 pixels images[1]. The values of alpha are taken from 0.05 to 0.5 and number of embeddings taken from 5 to 15. For measuring the imperceptibility we make use of the measure PSNR defined as follows:

$$PSNR = 10 \log_{10} (255^2 / MSE),$$

$$MSE=(1/N)^2 \sum\sum (x_{ij} – x'_{ij})^2 ,$$

where x denotes the original pixel value, and x' denotes the decoded pixel value.

Some of the results are summarized beow in the table 4.1 and figure 4.1.

5. CONCLUSION

We observe that choosing the value of alpha between 0 and 1, preferably 0.05, we can achieve best perceptibility. We also observe that the PSNR values decrease as we increase the number of times of embedding of message in the cover image, but still remains in the acceptable range of 35 to 40.

Our algorithm provides maximum embedded capacity in the cover image. The embedding capacity is equal to 3 times the number of pixels contained in the color image, i.e., capacity percentage is 100%.

There are multi-level securities proposed in our algorithm. We have encoded the message using self-synchronizing T-codes with a key and the encoded message is embedded in the Haar wavelet transform coefficients of image using another key, called stego_key. The Wavelet-fusion-technique further uses a value of alpha. This value is secret and shared by the sender and receiver. The value of alpha is used to adjust the normalized cover's pixels. The advantage of Best T-codes is seen at the decoding stage where due to its self-synchronizing property we obtain the original message even after if signal processing noise being added to stego_image.

6. NOISE ANALSIS

We have analyzed our algorithm for robustness by adding noise to stego-images of .jpg format. The results of their

[1] 'jpg'

PSNR so obtained are summarized in table 6.1 and figure 6.1.

ACKNOWLEDGMENT
The authors wish to thank their students Amogh Batra and Pankaj Malhotra for their support in implementing the algorithm in MATLAB.

REFERENCES

[1]. Anderson, R.J. and F.A.P. Petitcolas , " *On the limits of steganography",* IEEE J. Selected Areas in Commun., 16: 4, 1998

[2]. Johnson, N.F. and S. Jajodia, " Exploring steganography: Seeing the unseen", IEEE Computer, 31,1998, 26-34.

[3]. Chin-Chen Chang, Tung-Shou, and Lou-Zo Chung, "A steganographic method based on JPEG and quantization table modification", Information Sciences 141 , 2002, 123-138

[4]. Fong A.C.M., Higgie G.R., Fong B, " Multimedia Application of Self-Synchronizing T-codes", Proc. IEEE Int. Conf. On IT: Coding and Computing, April 2001, pp.519-523.

[5]. Fong A.C.M., Higgie G.R., " Using a tree algorithm to determine the average synchronization delay of self-synchronizingT-codes", IEE Proceedings:Computer Digit., Tech., Vol. 43, No. 3, May 2002, 79-81

[6]. Higgie G.R., " Analysis of the Families of Variable-Length Self-Synchronizing Codes called T-codes", PhD thesis, The University of Auckland, 1991.

[7]. Higgie G.R., " Database of best T-codes", IEE Proc. Comput. & Digital Techniques, Vol. 143, 1996, pp. 213-218.

[8]. Hong G.Y., Fong B, Fong A.C.M., " Error Localization for robust Video Transmission", IEEE Transaction on Comuter Electronics, Vol. 48, No.3, August 2002 , pp.463-469.

[9]. Stefan Katzenbeisser, Fabien A.P. Petitcolas:, "Information Hiding Techniques for Steganography and Digital Watermarking", Artech House,Inc, 2000

[10]. Mehdi Kharrazi, Husrev T. Sencar, and Nasir Memon:, "Image Steganography: Concepts and Practice", WSPC/Lecture Notes Series, April, 2004

[11]. Muttoo S.K. and Sushil Kumar, "*Robust Steganography using T-codes"*, IndiaCom 2007, Proceeding of the National Conference on Computing for Nation Development, BVICAM, New Delhi, 2007, pp. 221-223.

[12]. Muttoo S.K. and Sushil Kumar, "Data hiding in JPEG images", BVICAM's International Journal of Information Technology(IJIT), New Delhi-63, 2007 .Website: www.bvicam.ac.in

[13]. Muttoo S.K. and Sushil Kumar, "*Image steganography using self-synchronizing variable codes"*, International confrerence on Quality, Reliability and Infocom technology, MCMILLAN India Ltd., 2007

[14]. Muttoo S.K. and Sushil Kumar*, "* High Capacity Image steganography using Best T-codes"*,* Isdiks07, 2007

[15]. Pal S.K., Saxena P.K., and Muttoo S.K., "*Designing Secure and Survivable Stegosystems*", Defence Science Journal, Vol. 56, No. 2, April 2006, pp. 239-250.

[16]. Perkins S, Smith D.H., and Ryley A, "Robust Data Compression: Conssitency checking in the synchronization of variable length codes", The Computer Journal, Vol. 47, No. 3, 2004, 309-319

[17]. Provos Niels., "Outguess-universal steganography", http://www.outguess.org.

[18]. Provos Niels and Honeyman, "Hide and Seek: an Introduction to Steganography" , IEEE Security and Privacy, May/June, 2003, 32-43.

[19]. Rabah Kefa, "Steganography-The Art of Hiding Data", Information Technology Journal 3 (3), 2004, 245-269.

[20]. Reddy P., "Error resilient image transmission using T-codes and edge-embedding", M.Sc. thesis, lane Department of Computer Science and Electrical Engineering, Morgantown, West Virginia, 2007.

[21]. Salomon David: Data Compression, Springer-Verlag, N.Y., second edition, 2000.

[22]. Silva Eric. A, Agaian Sos. S, "The best transform in the replacement coefficients and the size of the payload relationship sense", IS&T Archiving Conference, April, San Antonio, 2004, 119-203

[23]. Titchener, M.R., " Technical note:Digital encoding by way of new T-codes", IEE Proc. E. Comput. Digit Tech, 131, (4), 1984, pp. 151-153.

[24]. Titchener, M.R., " Generalised T-codes: extended construction algorithm for self- synchronization codes", IEE Proc. Commun., Vol. 143, No.3, 1996, pp. 122-128.

[25]. Titchener, M.R., " The synchronization of variable length codes", IEEE Transaction of Information theory, Vo. 43, No. 2, March 1997, 683-691.

[26]. Tolba, M.F.; Ghonemy, M.A.-S.; Taha, I.A.-H.; Khalifa, A.S., "High Capacity Image Steganography using Wavelet-Based Fusion*"* ,Computers and Communications, 2004. Proceedings. ISCC 2004. Ninth International Symposium on Volume 1, Issue , 28 June-1 July 2004 ,Vol.1, 430-435

[27]. Gunther Ulrich,"Robust Source Coding with Generalised T-codes"*, a* thesis submitted in the University of Auckland, 1998

[28]. Upham D.: Jpeg-Jsteg, ftp://ftp.funet.fi/pub/crypt/steganography/jpeg-jsteg-v4.diff.gz.

[29]. Huaiqing Wang and shuozhong: Cyber warfare:steganography vs. steganalysis,

Communications of the ACM, Volume 47, Number 10, 2004.

[30]. A. Westfeld And A. Pfitzmann, "Attacks On Steganogrphic Systems", 3rd International Workshop On Information Hiding, 1999.

Alpha	num	PSNR	message received	Imperceptibility
0.05	5	45.8007	Y	Good
0.05	10	42.7976	Y	Good
0.05	15	40.9844	Y	Good
0.0^{2}7	15	38.1697	Y	Good
0.08	10	38.3816	Y	Good
0.09	5	39.9127	Y	Good
0.09	8	37.9525	Y	Good
0.09	10	37.0252	Y	Good
0.09	15	35.4170	Y	Good
0.1	10	35.9868	Y	Good
0.1	15	34.3201	Y	Good
0.15	10	35.9868	Y	Fair
0.25	5	38.7904	Y	Poor
0.25	10	35.9868	Y	Poor
0.5	5	38.7904	N	Zero

Table 4.1: Image : lena.jpg ; Embedding message length = 2734

Images	PSNR without noise	PSNR After Salt&Pepper	PSNR After Gaussian	Number Of Embeddings
I1	42.7976	34.1478	36.8520	3
I2	42.1442	33.6824	36.5151	3
I3	44.9595	33.8505	37.7551	3
I1	38.6900	32.4668	34.1898	8
I2	38.0618	31.9876	33.6815	8
I3	40.7435	32.8145	35.5421	8

Table 6.1: alpha = 0.07; I1='jaan'jpg'; I2='lena.jpg'; I3='Tulips.jpg'

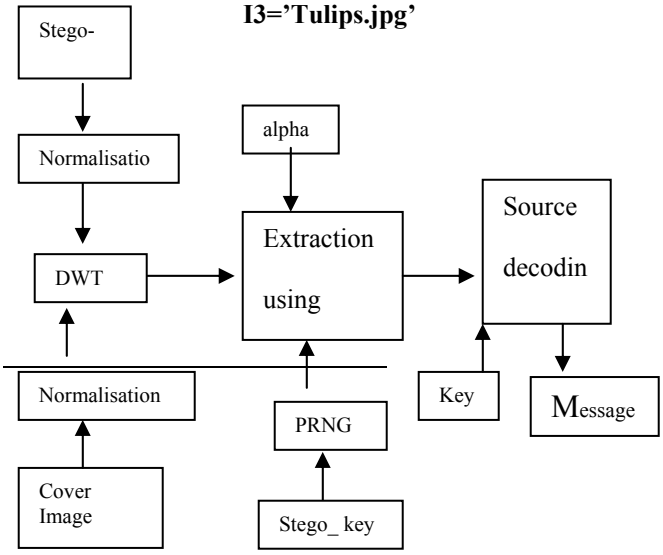


Figure 3.2: The block diagram of the message extraction

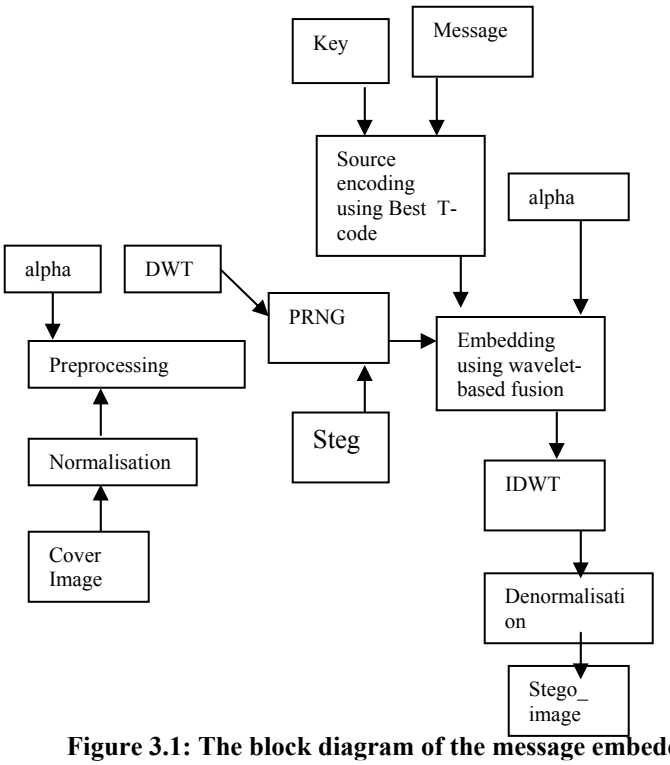


Figure 3.1: The block diagram of the message embedding

Figure 4.1: Stego_images of lena.jpg (in black and white)

(a)

(b)


(c)

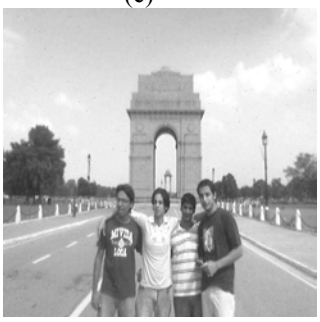

(d)


(k)


(l)


(e)


(f)


(g)


(h)


(i)

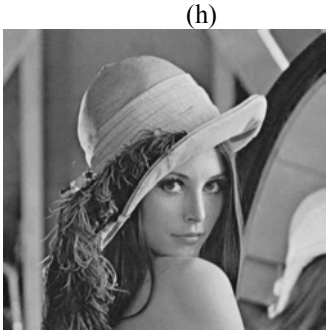

(j)

Figure 6.1: (a), (c), (e): after adding Salt & Pepper & noofembedding=3; (b), (d), (f): after adding Gaussian & noofembedding=3; (g), (i), (k): after adding Salt & Pepper & noofembedding=8; (h), (j), (l): after adding Gaussian & noofembedding=8;