TOWARDS OBJECTIFYING INFORMATION HIDING

Abbas Cheddad, Joan Condell, Kevin Curran and Paul Mc Kevitt

School of Computing and Intelligent Systems, Faculty of Computing and Engineering,
University of Ulster, Derry, BT48 7JL. E-mail: cheddad-a@email.ulster.ac.uk

ABSTRACT

In this paper, the concept of object-oriented embedding (OOE) is introduced into information hiding in general and particularly to steganography which is the science that involves undetectable communication of secret data in an appropriate multimedia carrier. The proposal takes advantage of computer vision to orient the embedding process. Although, any existing algorithm can benefit from this technique to enhance its performance against steganalysis attacks, however this work also considers a new embedding algorithm in the wavelet domain using the Binary Reflected Gray Code (BRGC). In the realm of information hiding, one wing focuses on robustness, i.e., watermarking, and another wing focuses on imperceptibility, i.e., steganography. This work advocates for a new steganographic model that meets both robustness as well as imperceptibility. Resilience against common steganalysis attacks including the 274-D merged Markov and DCT features while surviving various image processing manipulations are reported. A neural network classifier was trained with features derived from 400 images. Comparisons with existing systems will also be highlighted.

KEYWORDS: Image processing, Digital communication, Wavelet transforms.

1. INTRODUCTION AND BACKGROUND

In the realm of this digital world, steganography has created an atmosphere of corporate vigilance that has spawned various interesting applications, thus its continuing evolution is guaranteed [1]. This work makes use of some terms commonly used by steganography and watermarking communities. Therefore the authors assume the reader is familiar with such terms.

To our knowledge no prior work has discussed the application of skin tone detection in conjunction with adaptive steganography. All of the prior steganographic methods suffer from intolerance to any kind of image manipulations applied to the stego-image, i.e., a Warden active attack scenario [2]. Scholars differ about the importance of robustness in steganography system design. In [3], Cox regards steganography as a process that should not consider robustness as it is then difficult to differentiate from watermarking. Katzenbeisser, on the other hand, dedicated a sub-section to robust steganography. He mentioned that robustness is a practical requirement for a steganography system. "Many steganography systems are designed to be robust against a specific class of mapping." [4]. It is also rational to create an undetectable steganography algorithm that is capable of resisting common image processing manipulations that might occur by accident and not necessarily via an attack.

Since in steganography the choice of cover images is not rigid, we decide to target images with human presence. The automatic detection of human skin tone is of utmost importance in numerous applications such as, video surveillance, face and gesture recognition, human computer interaction, image and video indexing and retrieval, image editing, vehicle drivers' drowsiness detection, controlling users' browsing behaviour. It has received considerable attention in recent years [5, 6], especially in the areas of biometrics and computer vision.

Various steganographic methods have been introduced. They can be categorized into three groups: spatial domain methods, frequency domain methods and adaptive methods or model based, which are essentially special cases of the former two. There exist hundreds of steganographic methods therefore we restrain our discussion to the most popular ones in each category.

In the spatial domain a steganographer modifies the secret data and the cover medium which involves encoding at the level of the LSBs (least significant bits). This method although simpler, has a larger impact compared to the other two types of methods [7]. S-Tools is a method that involves changing the least significant bit of each of the three colours in a pixel in a 24-bit image, for example a 24-bit BMP file [8, 9]. The problem with 24-bit BMP images is that they are not commonly used on the web and tend to stand out (unlike JPEG and PNG). S-Tools, as well as other tools based on LSBs in the spatial domain, take for granted that least significant bits of image data are uncorrelated noise [10].

F5 was the creation of Andreas Westfeld in 2001 [9]; it embeds messages by modifying the DCT (discrete cosine transform) coefficients. The central operation done by F5 is matrix embedding (subtraction and matrix encoding) with the aim of reducing the amount of changes made to the DCT coefficients. Abdulaziz and Pang [11] use vector quantization called Linde-Buzo-Gray (LBG) coupled with Block codes known as BCH code and 1-Stage discrete Haar Wavelet transforms. They reaffirm that modifying data using a wavelet transformation preserves good quality with little perceptual artefacts. Abdelwahab and Hassan [12] propose a data hiding technique in the DWT domain. Both secret and cover images are decomposed using DWT (first level), each of which is divided into disjoint 4x4 blocks. Blocks of the secret image are fitted into the cover blocks to determine the best match. Afterwards, error blocks are generated and embedded into coefficients of the best matched blocks in the HL of the cover image. Note that the extracted payload is not totally identical to the embedded version.

Adaptive steganography is a special case of the two former methods. It is also known as "Statistics-aware embedding" [13], "Masking" [14], "Model-Based" method [15] and "block complexity" [16]. This method takes statistical global features of the image before attempting to interact with its LSB/DCT coefficients. The statistics will dictate where to make the changes [17, 18]. It is characterized by a random adaptive selection of pixels depending on the cover image and the selection of pixels in a block with large local STD (standard deviation). The latter is meant to avoid areas of uniform colour, e.g., smooth areas. This behaviour makes adaptive steganography seek images with existing

or deliberately added noise and images that demonstrate colour complexity.

Spatial domain algorithms are prone to statistical attacks such as Chi-Square χ^2 [19] and ± 1 steganalysis [20]. Frequency domain, i.e., JPEG, methods are venerable to attacks in the form of double compression effect, statistical distribution of DCT coefficients and merged statistical features as we will discuss later.

This paper is organised as follows. A brief disclosure of two algorithms to segment regions of interest in grayscale and colour images is discussed in Section 2.1. In Section 2.2, a wavelet steganographic algorithm is provided. Then Section 3 emphasizes the robustness of the algorithm. Section 4 highlights the limitation and merits of this work. Finally, the conclusion is drawn in Section 5.

2. THE PROPOSED SCHEME

2.1. Image content segmentation

For grayscale face images, we use the algorithm described in [21], which has the advantage of extreme ease of implementation. Given a set of 2D points, the Voronoi region for a point P_i is defined as the set of all the points that are closer to P_i than to any other points. More formally we can say: Let $S = \{P_1, P_2, \ldots P_n\}$ be a finite subset of R_m and let $d: R_m \times R_m \rightarrow R$ be a metric. We define the Voronoi region $VR(P_i)$ of a point P_i via $VR(P_i) = \{P \in R_m \mid d(P, P_i) \leq d(P, P_j)$ for all $j = 1, 2, \ldots, n, j \neq i\}$, i.e., $VR(P_i)$ is the set of all points that are at least as close to Pi as to any other point of S. The set of all 'n' VR is called the Voronoi Diagram $VD(S)$ of S [22]. VD is generated from a set of sites that correspond to the image histogram bin values. In essence, these points are ≤ 255. A set of triangulation vertices is then produced (Delaunay Triangulation) which dictates the graph cut. After image segmentation template matching is used to vote for a face blob.

For colour face images, we use the algorithm described in [23], a skin probability map is created from a special non-linear transformation that injects a zeroed R (the red component in RGB images) into its formulation.

2.2. The proposed embedding algorithm

The central focus of this paper is to embed the secret message into the approximation decomposition in the first-level 2D Haar DWT with the symmetric-padding mode guided by the detected skin tone areas. A coefficient's precision is left intact while only its integer element carries the secret bit using BRGC. The last LSB where the steg-value, compared to the plain-value, is unchanged, increased or decreased by one (i.e., change by ± 1 in the 1st LSB or ± 4 in the 3rd LSB) eventually leaves traceable statistical violations. Many algorithms to date still use such conventional models either in the spatial domain or the transform domain. The BRGC allows alteration to even the third LSB (i.e., change by ± 3) in the DWT without much degradation compared to the conventional use of PBC (Pure Binary Code). Let a plain-image pixel at the approximation level of a 1st level DWT be the coefficient C and let the secret bit be '0': C=325.09821988712.

- BRGC

C_{int}=325, Store=.09821988712
BRGC (C_{int}) = '111100111'
Steg-image (BRGC (C_{int})) ='111100011'
BRGC -to-Decimal='111100011'➔322
Steg-image=Concatenate (322, Store) = 322.09821988712
Difference ± 3 (odd number).

- PBC

Bin (C_{int}) = $(101000101)_2$
Steg-image (Bin (C_{int})) = $(101000001)_2$
Bin-to-Decimal = $(101000001)_2$➔321
Steg-image= Concatenate (321, Store) = 321. 09821988712
Difference ± 4 (even number).

The algorithm starts by first segmenting probable human skin regions, using the former mentioned methods, such that:

$$C = C_{bg} \cup C_{fg}, \text{where}, C_{fg} \in \{\bigcup_{i=1}^{n} S_i\}, S_i \cap S_j = \varnothing, \forall i \neq j \quad (1)$$

In Eq. (1) C, Cbg, and Cfg denote the cover image, the background regions and the foreground regions respectively. $\varnothing$ denotes the empty set and (S_1, S_2,…, S_n) are connected subsets that correspond to skin regions. We found that embedding into these regions produces less distortion to the carrier image compared to embedding in a sequential order or in any other areas. Such phenomena result from the fact that the eye does not respond with equal weight of sensitivity to all visual information. This is consistent with the claim that certain information simply has less relative importance than other information in the human visual system. This information is said to be psycho-visually redundant since it can be altered without significantly impairing the quality of the image perception [24]. Human presence in digital photography and video files encourages such an approach. In this context, the postulation of the above skin model would definitely help in the case of image translation as it is invariant to such distortions. With reference to Eq. 2, if the cover image is geometrically transformed by a translation of tx, along the x axis, and ty, along the y axis, in such a way that the new coordinates are given by:

$$C\begin{bmatrix} x' \\ y' \end{bmatrix} = C\begin{bmatrix} x + t_x \\ y + t_y \end{bmatrix} \quad (2)$$

then each detected skin blob will be transformed likewise with the same distance to the origin as shown in Eq. 3. Skin regions are extracted based on colour tone; and hence, are undisturbed by translation.

$$S_i\begin{bmatrix} x' \\ y' \end{bmatrix} = S_i\begin{bmatrix} x + t_x \\ y + t_y \end{bmatrix}, \forall i \in \{1,\ldots,n\} \quad (3)$$

To cope with rotation, it is sufficient to locate face features, i.e., eyes, based on the method described in [21]. Salient features form reference points that dictate the orientation of embedding and thus aid recovery from rotational distortions. Fig. 1 (left) shows an attacked stego image with joint attacks of cropping, JPEG compression, translation (offset of 60 pixels) and rotation (-30 degrees) - shown with the extracted secret data and (right) attacked with salt and pepper noise- shown along with extracted secret data.

Fig. 1. Resistance to image processing attacks.

Rotation about the origin is defined as in Eq.4.

$$x' = x\cos\theta - y\sin\theta, y' = x\sin\theta + y\cos\theta \quad (4)$$

The angle θ will be determined from the elliptical model formed by face blob. If the attacked image is rotated in the opposite direction with the same angle, i.e., $\theta' = -\Delta\theta$ caused by the attack, the method will be able to restore the angle and will have the coordinates as shown in Eq.5.

$$x' = x, y' = y \quad (5)$$

Eq. 5 is used where embedding occurs in the neutralised orientation where $b_{axis} \perp x_{baseline}$ (b_{axis} denotes the major axis in a face ellipse). However, the encoder has 359 choices for the angle as expressed in Eq. 6.

$$\theta' = \theta \pm \alpha \quad (6)$$

where $\alpha \in \{1,2,...,359^{\circ}\}$ denotes an agreed upon scalar which can form another optional secret key as shown in Fig 2. Note that, for simplicity, α here belongs to the discrete space while in practice is continuous. However, the use of discrete values is encouraged in order to minimise the errors in the recovered bits.

Fig. 2. (left) stego wrongly de-rotated to $\theta = -183$ and the retrieved data, (right) stego correctly de-rotated to $\theta = -184$ and the retrieved data.

In addition to this, the algorithm yields a robust output against reasonable noise attacks and translation. Robustness against noise is due to the embedding in the 1st-level 2D Haar DWT (Discrete Wavelet Transform).

Algorithms based on DWT experience some data loss since the reverse transform truncates the values if they go beyond the lower and upper boundaries, i.e., 0-255. Knowing that human skin tone resides along the middle range allows us to embed in the DWT without worrying about the truncation. In the case of colour images, this would leave the perceptibility of the stego-image virtually unchanged since the changes made in one component of the YC_bC_r will be spread evenly among the *RGB* colours when transformed back. We choose wavelets over DCT (Discrete Cosine Transform) because: the wavelet transform mimics the Human Vision System (HVS) more closely than DCT does; visual artefacts introduced by wavelets coded images are less evident compared to DCT because the wavelets transform does not decompose the image into blocks for processing.

3. STEGANALYSIS AND VISUAL PERCEPTIBILITY

In the frequency domain, Pevny and Fridrich [25] developed a multi-class JPEG steganalysis system that comprises of DCT features and calibrated Markov features, which were then merged to produce a 274-dimensional feature vector. This vector is fed into a Support Vector Machine (SVM) multi-classifier capable of detecting the presence of Model-Based steganography, F5, OutGuess, Steghide and JP Hide&Seek.

Initially, we thought of reducing the complexity of the 274-D vector by retaining only the most contributing features using Principal Component Analysis (PCA) but opt not to go for that as we found a comment against such procedure in [26]. We created features derived from 200 images demonstrating different structural complexities obtained using various digital camera models in addition to images downloaded from the Internet. We generated another 200 stego-images using the same set and similarly obtained their features. Then we created a feed-forward back-propagation network instead of SVM to act as a classifier and we fed into it the 400 feature vectors. An independent testing set comprising 80 images was used to simulate the network. The result confirms that the proposed scheme can overcome detection using this attack. A surprising observation was that the detection rate was slightly better when the payload was small unlike when the full skin area was used. The reported detection probability is still within a random guessing range: sim(net,Set_Small) => 36.8421%, sim(net,Set_Full)=> 31.5789%.

The second attack, namely ± 1 steganalysis, can not be accomplished since the embedding changes do not produce this effect, see Fig 3 which contrasts our algorithm to S-Tools which is prone to this kind of attacks.

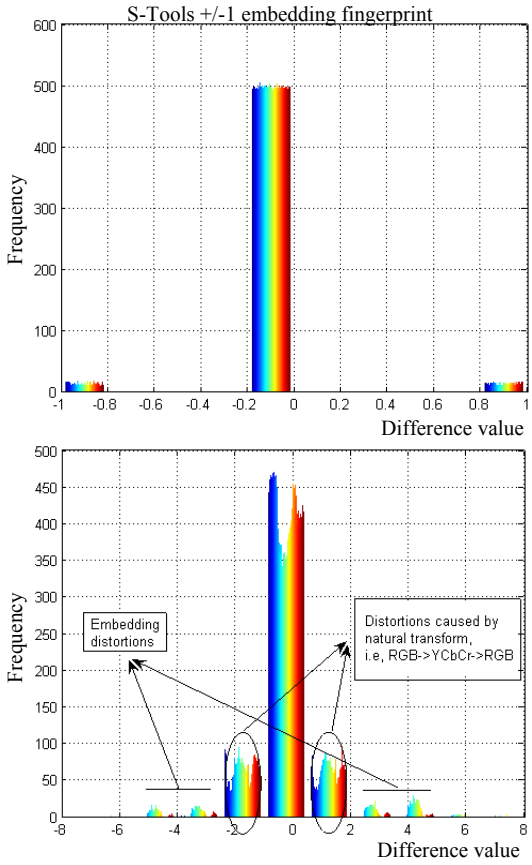

Fig. 3. Changes after embedding, (top) S-Tools' ± 1 embedding fingerprint (bottom) the proposed method.

4. LIMITATIONS AND MERITS

The first thing that comes to one's mind is the limited payload available by targeting skin regions. Extending this method to video files would be a possible remedy. However, a steganographer may choose to consider the entire image for embedding, and then detecting skin area would reduce to just providing the desired

secret embedding angle. Targeted embedding methods, such as the new enhanced MB2, are faced with much more accurate targeted attacks. That is because "if the selection channel is public, the attacker can focus on areas that were likely modified and use those less likely to have been modified for comparison/calibration purposes" [27]. Nevertheless, the proposed scheme has some advantages. Choosing a specific secret embedding angle would help existing attacked algorithms fool steganalysis tests. Moreover, when an image is de-rotated to its pristine angle state, interpolation occurs ($\theta \notin \{0,90,180,270,360°\}$) offering a practical method for minimizing embedding impact. Identifying skin areas will give an instant direct split of an image into two main areas, one for embedding and another to correct for any statistical distortion caused.

5. CONCLUSION

This paper reaffirms that an object-oriented embedding approach to steganography is possible thanks to established computer vision algorithms. The proposed application is designed in order to serve as a test-bed for our ongoing research on enhancing steganography in digital images.

6. REFERENCES

[1] G.J. Simmons, "The prisoners' problem and the subliminal channel," in Proc. of Advances in Cryptology, pp. 51-67, 22-24 August. 1984.

[2] L. Siwei, "Natural Image Statistics for Digital Image Forensics," Thesis of Doctor of Philosophy, Dartmouth College, Hanover, New Hampshire, pp. 67, August, 2005.

[3] I. Cox, Information hiding, watermarking and steganography, Public Seminar, Intelligent Systems Research Centre, University of Ulster, Northern Ireland, 28th April 2009.

[4] S.C. Katzenbeisser, Principles of steganography, In: S. Katzenbeisser and F.A.P Petitcolas, (ed.), *Information hiding techniques for steganography and digital watermarking*, Norwood: Artech House, INC, 2000.

[5] M. Corey, F. Farzam and C.J. Herng, "The effect of linearization of range in skin detection," in Proc. IEEE 6th International Conference on Information, Communications & Signal Processing, pp.1-5, 10-13 Dec. 2007.

[6] U.A. Khan, M.I. Cheema and N.M. Sheikh, "Adaptive video encoding based on skin tone region detection," in Proc. IEEE Students Conference Proceedings, 16-17 Aug. 2002, vol. 1, pp.129-34.

[7] P. Alvarez, "Using extended file information (EXIF) file headers in digital evidence analysis," *International Journal of Digital Evidence*, vol. 2, no. 3, winter 2004.

[8] P. Wayner, *Disappearing Cryptography*, 2nd Ed. USA, Morgan Kaufmann Publishers, 2002.

[9] N.F. Johnson, D. Zoran and J. Sushil, *Information Hiding Steganography and Watermarking - Attacks and Countermeasures*, Kluwer Academic Publishers, 2001.

[10] A. Westfield and A. Pfitzmann, "Attacks on steganographic systems breaking the steganography utilities EzStego, Jsteg, Steganos and S-Tools and some lessons learned," in Proc. 3rd International Workshop on Information Hiding, Dresden Germany, LNCS 1768, pp. 61-76, September/October 1999.

[11] N.K. Abdulaziz and K.K. Pang, "Robust data hiding for images," in Proc. IEEE International Conference on Communication Technology, vol. 1, pp. 380-383, 21-25 Aug. 2000.

[12] A. A. Abdelwahab and L.A. Hassan, "A discrete wavelet transform based technique for image data hiding," in Proc. 25th National Radio Science Conference, Egypt, pp.1-9, 18-20 March. 2008.

[13] N. Provos and P. Honeyman, "Hide and seek: An introduction to steganography," *IEEE Security and Privacy*, vol.1, no. 3, pp.32-44, May-June 2003.

[14] N.F. Johnson and S. Jajodia, "Exploring steganography: Seeing the unseen," *IEEE Computer*, vol. 31, no. 2, pp.26-34, Feb. 1998.

[15] P. Sallee, "Model-based methods for steganography and steganalysis," *International Journal of Image and graphics*, vol. 5, no. 1, pp. 167-190, 2005.

[16] H. Hioki, "A data embedding method using BPCS principle with new complexity measures," in Proc. Pacific Rim Workshop on Digital Steganography, pp.30-47, 2002.

[17] R. Tzschoppe, R. Baum, J. Huber and A. Kaup, "Steganographic system based on higher-order statistics," in Proc. SPIE, Security and Watermarking of Multimedia Contents V, Santa Clara, California, USA, vol. 5020, pp. 156-166, 2003.

[18] E. Franz, "Steganography preserving statistical properties," in Proc. of the 5th International Workshop on Information Hiding, Noordwijkerhout, The Netherlands, *LNCS*, vol. 2578/2003, pp. 278-294, 2003.

[19] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems," in Proc. of the 3rd Workshop on Information Hiding, Dresden, Germany, LNCS 1768, pp. 61-76 , 2000.

[20] G. Cancelli, G. J. Doërr, M. Barni and I.J. Cox, "A comparative study of ±1 steganalyzers," in Proc. of IEEE 10th Workshop on Multimedia Signal Processing, pp.791-796, 8-10 Oct. 2008.

[21] A. Cheddad, D. Mohamad and A. Abd Manaf, "Exploiting Voronoi diagram properties in face segmentation and features extraction," *Pattern Recognition*, vol. 41, no.12, pp.3842-3859, 2008.

[22] L. Costa, R. Cesar, *Shape Analysis and Classification*, CRC Press, USA, 2001.

[23] A. Cheddad, J. Condell, K. Curran and P. Mc Kevitt, "A skin tone detection algorithm for an adaptive approach to steganography," *Signal Processing*, 89(12)(2009) 2465-2478.

[24] R. C. Gonzalez and R.E. Woods, *Digital Image Processing*, Prentice Hall, ch. 8, pp. 417, 2002.

[25] T. Pevnya and J. Fridrich, "Merging Markov and DCT features for multi-class JPEG steganalysis," in Proc. of SPIE Electronic Imaging, Photonics West, pp. 03-04, January 2007.

[26] J. Kodovský and J. Fridrich, "On completeness of feature spaces in steganalysis," in Proc. of the 10th ACM workshop on Multimedia and security, Oxford, UK, pp. 123-132, Sept. 22-23, 2008.

[27] J. Kodovský and J. Fridrich, "Influence of embedding strategies on security of steganographic methods in the JPEG domain," in Proc. of SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, San Jose, CA, vol. 6819, pp.1-13, January 28-31, 2008.