2009 International Conference on Machine Learning and Computing
IPCSIT vol.3 (2011) © (2011) IACSIT Press, Singapore

Novel Spatial Domain Reversible Information Hiding Technique for Content Authentication

Fahad Ali Usmani, Ekram Khan

Department of Electronics Engineering, AMU, Aligarh, India

Abstract. In this paper we propose a transparent, computationally simple and fast spatial domain data hiding technique for content (image and audio) authentication purposes. It can also be used in data captioning, copyright protection, confidentiality and integrity over the property distributed in digital formats over various transmission channels but its fragile nature prevents it to deliver efficient performance in such robust applications. Our method is based on concept that an image (host) can be transformed to another known image (target) by embedding controlled amount of distortion at that particular pixel value controlled through payload bits. Results obtained using MATLAB show that for lower distortion levels, PSNR values as high as 53dB while for appreciable distortions, not below the 35dB level are easily achievable. For our case, the capacity is image and distortion control parameter dependent, hence ranges from 500 bits to 33Kb for a 512x512x8 image requiring a computation time of 105 ms. Moreover, the same technique has also been applied to audio data and even better results obtained are reported. The attractive feature of the proposed algorithm is its simplicity in both embedding and detection of the payload bits, computational efficiency and apparent nature of marked information.

Keywords: Data Hiding, distortion, fragility, robustness, transparency

1. Introduction

Increasing advancements in communication technology for data transfer over Internet and high speed digital subscriber lines introduce new challenges and problems regarding digital data security. When on one hand, digital media proves to be a convenient source of data transmission then on the other it has its own shortcomings in providing security to content receivers. One of the most important problems is to prevent unauthorized copying i.e. duplication and redistribution of digital productions that too without any loss of quality, may it be data, images, audio and video. Digital watermarking techniques have proved to be the powerful way to claim intellectual protection and signal certification [1]. The ease with which the earlier algorithms were broken has given rise to the development of newer information hiding and security providing algorithms.

Data hiding is a general process which links two set of data i.e. embedded data and cover media data, both of which require the imperceptibility of embedded data [2]. Specifically for content authentication purposes, embedded data should be closely related to the cover media and most of the times retrieval of original image is difficult when received data have been corrupted through noisy channel [3]. Reversible data hiding facilitates immense possibility of applications to link two set of data to perfectly recover the originally transmitted signal. For copyright protection purposes, any data hiding algorithm must at least meet the preliminary requirements of Transparency, Robustness and Security [4]. There exist a trade-off between the transparency and the robustness of the data hiding technique used, which is always critical in any particular application where resolving one parameter is to trade-off the other. Being the most obvious solution for all kind of fraud and unlawful impersonation and software piracy, information hiding techniques are better suited because the security is essentially tied to intellectual property: the actual content (opposed to the digital representation) is protected. It is also of interest in non-security applications such as in commerce and

in focus, it has gained interest in applications involving captioning, maintenance of audit trails, and embedding hyperlink Information in “hyper-media.”, temper assessment/authentications, cover communication and trace applications etc [5-6].

In the past many efforts have already been applied in this direction but most of them being irreversible e.g. widely used spread spectrum based hiding models [7-8], LSB based schemes which are in fact lossy [9], Quantization index modulation (QIM) based schemes suffering from quantization error problems etc. [10] More recently, some reversible data hiding schemes have emerged such as spatial domain technique [11] using modulo-256 addition to embed the hash value of the original image for authentication and [3] utilizing the zero or minimum points of the histogram and modifying the pixel grayscale values, transform domain marking technique [12] using a lossless multiresolution transform and idea of a patchwork having very limited applications and most popular scheme capable of embedding large amount of data [13] using image segmentation and compression techniques but suffering from limited payload.

Figure1.Dan (a) Host image(X) (b) Target image(Y) (c) Marked (W) image

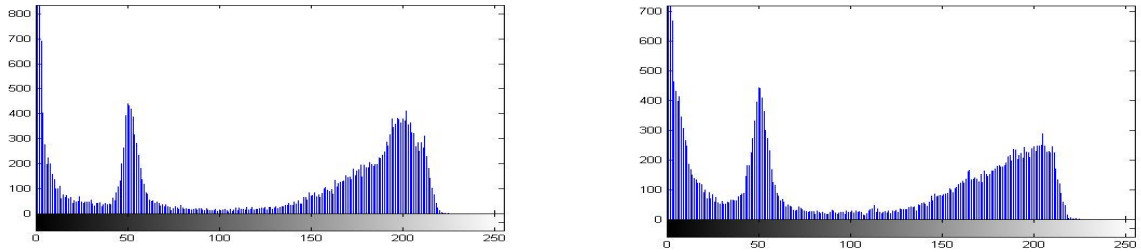

Figure2.Histogram for (a) Host Image ‘X’ and (b) Target Image ‘Y’

2. Overview of proposed Scheme

In this spatial domain data embedding scheme, we propose an algorithm, simpler to implement and powerful in disillusioning the pirates and attackers from getting the wrong perception of what have been transmitted along with the general message and requires the original image (host or target) for the detection of its credibility. In general, when a stream of data passes by the user, he checks it by the most common sense available to him and guesses it to be the one as a message with no information due to his inability to detect whether the message in hand is the one that correspond to the original or that have been encrypted to make it look like another similar information that is made to resemble much with the one made the ‘target’. Payload bits are embedded through image conversion, which we name as ‘image riding’ as we purposefully try to override one image over the other. This method is successful in tackling the situation from the problem point of view as we try to insert distortion deliberately and exploit for information hiding purposes.

2.1. Information embedding Algorithm

The proposed information hiding scheme is described in the form of an algorithm below that demarcates the effect of embedding the information over the host image with the motive of its subsequent conversion to the ‘target image’ (one which resembles much with the host image and serves as a target for the addition of the distortion in the message (host)), a general example of Dan as shown in figure 1(a, b).The two images belong to the same person but taken over a span of 3 years (or any other criteria may be chosen).

For a given greyscale bitmap image (Let us say, Dan’s image of size 160 X 200 X 8) with each pixel greyscale value variable from 0 to 255, we proceeds through the steps listed below:

1. Get the ‘Information payload matrix, Q’, ‘host image, X’ and ‘target image, Y’. Choose initially the level of distortion to be added, denoted by ‘n’ called distortion control parameter.

2. Compare X with Y on the basis of ‘n’ and generate a new same sized matrix ‘u’ after comparison using criterion given below.
3. If Y(i,j)-X(i,j)>=n, then u(i,j)=’1’, else if X(i,j)-Y(i,j)>=n then u(i,j)=’-1’ and for all other options, set u(i,j)=’0’
4. Load bit by bit the Payload stream as a count, ‘Q’ and use u(i,j) obtained above to embed bits sequentially in X to generate a new matrix ‘W’ using criterion given below. This complex step is necessary to ensure the correct embedded information extraction (detection) at the receiver output.
5. If Q(i,j)=’1’ then get next pixel location with entry ‘1’ in u(i,j) and add ‘n’ to the corresponding pixel location in X and if Q(i,j)=’0’ then subtract n.
6. If count =0 then stop the algorithm and obtain a new image out of X called W as shown in Figure 2(c) for n=30 and Q= 757. Calculate processing time.

Attraction for the work followed comes from the slight difference in pixel intensity values appearing in the histogram representation of images X and Y. Figure 2 shows that to make one image resemble the other, it is possible to add ‘n’ at some pixel location while subtracting it from the other and through this method, addition and subtraction may simply be controlled by a random binary sequence (information, Q).

2.2. Information extraction Algorithm

After information embedding, the marked data is finally received where it is required to decode the message attached for the verification of authorised copy. The detection algorithm is described below:

1. Get the Marked image, W from the receiver and obtain the host image, X and start scanning and comparing both the images to obtain a new matrix ‘v’ consisting of entries ‘n’, ‘-n’ and ‘0’.
2. Count above ‘n’ and ‘-n’ entries and define a vector Q’ of above count. Proceed ahead following the sequence given below.
3. If v(i, j)=’n’ then get next entry in Q’ as ‘1’ otherwise if v(i, j)=’-n’ then get next entry in Q’ as ‘0’. For all other cases, do nothing.
4. Hence, we are able to decode the data embedded in W to get the information vector C which ideally should match with Q but due to various attacks and channel actions, it will differ thereby defining the robustness of the technique adopted.

In this way all the entries so obtained are listed in the form of a binary random code, Q’, the extracted information vector which ideally should match with the transmitted Q. This transmission accuracy depends on the robustness and transparency of the marking technique adopted.

3. Experimental results and discussion

To demonstrate the effectiveness of the proposed algorithm, different image processing and geometric distortion operations have been applied using Matlab 7.6.0.324 and Turbo C and the results have been reported to show the authenticity of the work. Since, here the energy of the cover signal is distributed over the entire image; this point enables application of this scheme in a better temper detection. Also the embedding of the information is dependent on the host image which decides the particular pixels free for access, hence putting the maximum limit on the payload bits that can successfully be embedded while keeping the integrity of the scheme intact as tabulated for ‘Dan’ image in Table 1.

Above picture of embedded bits shows that with increased distortion, the embeddable payload count decreases, leading to a sort of game with the distortion count. The effectiveness of work becomes clear from the peak signal to noise ratio (PSNR) given by Eq. 1 which is taken acceptable for values above 35dB which here comes out to be much better as depicted in fig. 3 (a) demarcating the effect of ‘n’ and ‘Q’ on the PSNR values..This actually means that even after the addition of distortion for embedding the information, noise performance of the algorithm is not seriously deteriorated. Moreover, we observe from Table I, that the capacity decreases for increasing values of distortion which is obvious and hence for lower distortions, capacity as high as 4030 Kb for (512, 512, 8) image are easily achievable. A good analysis of similarity is made through the calculation of correlation coefficient between the two host (X), marked (W) and target (Y) images showing the variation of the similarity through comparison as shown in Fig 3 (b).

$$PSNR = 10\log_{10}\left(\frac{160 \times 200}{MSE}\right) \quad \textbf{(1)}$$

A limiting feature of this scheme is the upper limit on payload bits to be used while success point is that whatever may be the message, it will create a new dependent watermarked image. We also find that the computation (embedding) time with a computer Pentium(R) D 2.80 GHz and the software MATLAB 7.6.0.324 is a total average of 33ms for embedding 4030 bits on a (160x200x8) image. Hence in different applications, same message could be transformed to any desired prototype and sent to different users, all having the original data effects of attacks on information detection. In practice, after the image is received, the information vector does not match with originally transmitted hidden information. Intentional attacks try

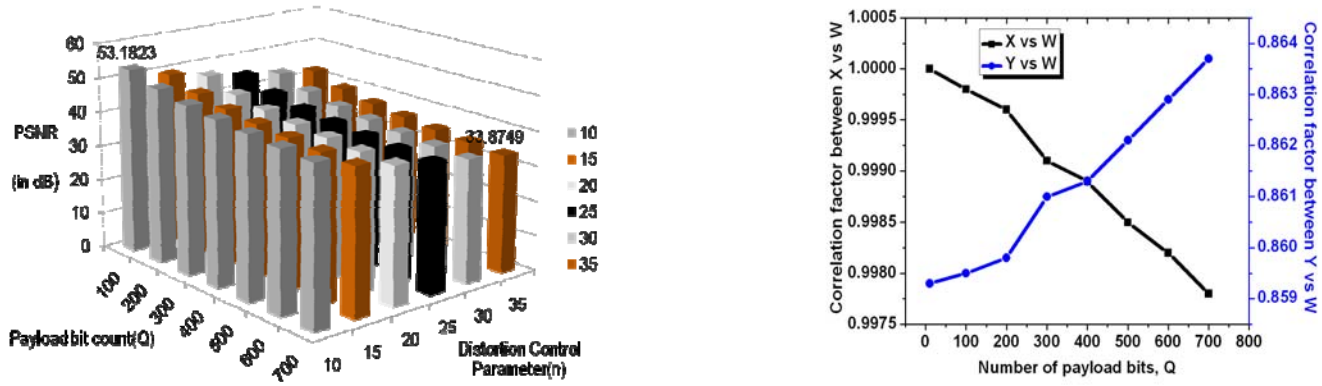

Figure3. Variation of (a) PSNR with 'n' and 'Q' (b) Correlation coefficient of marked image with both host and target

Figure4.Demonstration of Robustness with input information length (Q) to be 733 and 'n' to be 30; (a) JPEG compressed (quality = 10) watermarked image; (b) Watermarked image by 3 × 3 median filtering (c) Gaussian Noisy watermarked image (0, 0.01); (d) Sharpened Watermarked image using filter parameter ones(7)/49 (e) Lowpass filtered Watermarked image using filter parameters ones(3)/9

to degrade the image to a usable entity but losing its hidden information to a greater extent while unintentional attacks may go to any limits. Some most popular attacks include compression and enhanced attack techniques like noise, filtering, and sharpening etc. We also observe that our technique is very less prone to random bit attacks as determined from the correlation coefficient factor between transmitted and retrieved information (CCF) which comes out to be 1.0. We discuss all of them with reference to our proposed scheme.

Table I. Variation of payload bits with distortion

Distortion control parameter(n)	Maximum embeddable bits(Capacity)	Computation time (ms)
10	1565	34
15	1207	33
20	999	32
25	834	34
30	727	33
35	654	31

Table II. Variation of capacity and computation time with distortion control parameter for audio signal samples

n	C	Corr (X,W)	Corr (Y,W)	Time (ms)
0.001	2566	1.0000	0.0152	132
0.005	2234	0.9998	0.0158	135
0.01	1993	0.9993	0.0168	134
0.05	1369	0.9887	0.0300	141
0.1	0885	0.9711	0.0498	127

1. *Compression:* It is an unintentional attack which is deliberately applied to save the bandwidth. Normally compression is lossy such as JPEG (quality factor of 10 considered here), for which CCF comes

out to be 0.2 as shown in Fig 4(a). We vary the quality of the compression technique and find that with improved quality, the retrieved information is correct to certain acceptable standards. We assume that the Input information (payload) is of length 10 and common for all attacks given by Q = [1 0 1 0 0 1 0 1 1 0 1].

2. *Filtering*: Low pass filtering has negligible effect on image degradation but can seriously affect the performance of the system in some extreme cases. In median filtering, we define a 'tuple' as [m n] which defines a neighbourhood of size m x n over which the median is computed. CCF comes out to be 0.2182 and 0.333 respectively for both kind of filtering as shown in fig 4 (b & e) respectively.

3. *Noise:* We examine the effect of Gaussian noise on the received information vector. With a noise addition of variance 0.01, the image is degraded as shown in fig 4(c). We also observe that with increasing noise density for salt and pepper noise case and variance in zero mean additive white Gaussian noise case, received information is degraded appreciably demarcating its fragile nature by observing the CCF to be -0.333.

4. *Sharpening:* Since in our case, marked image does depend on information, so statistical averaging could easily be avoided. Moreover, sharpening effects also introduce certain distortions in the transmitted information such as CCF obtained in this case is 0.2182 as shown in Fig 4(d).

4. Extension of proposed technique to audio

To cover the 1-D audio domain we chose to record an audio track containing 50,000 samples sampled at 44.1 KHz as a mono audio data samples as shown in Fig. 5 and the obtained results are summarized in Table II. We find that here also the capacity depends upon the distortion control parameter, 'n' and type of audio signal chosen for comparison. Maximum embeddable capacity is found out to be around 22 Kb with an average processing time of 130ms (larger compared to previous case of image having 32k samples only, while here 50k). Authenticity of the technique applied can once again has been tested by observing the correlation factor values between original host and target signals with marked image, W. Similar to the embedding process, extraction have also been carried out as performed for the image case. We notice that the distortion caused by filtering and Gaussian noise addition is less pronounced here as compared to images and hence the signal is affected very little. This offers us a good option for using our technique in a wide range of audio signal transmissions over digital transmission channels.

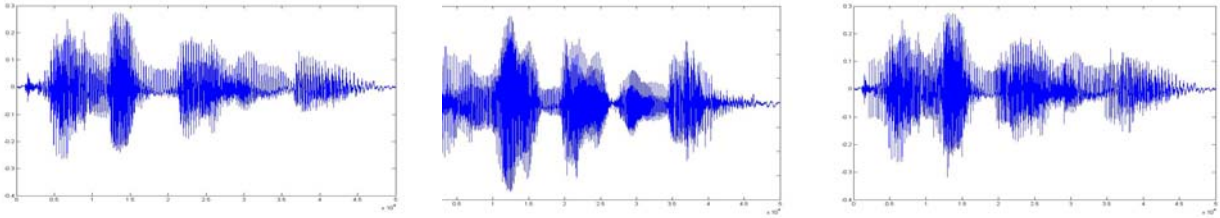

Figure4.(a) Sample plot of Audio Sample I (b) Sample plot of Audio sample II (c) Sample plot of marked audio data

5. Conclusions and Future scope

In this paper we have proposed a simple, transparent but fragile technique for hiding information in images and audio. It is based on the concept that the host image can be transformed to another similar image (target image) by controlling the distortion through information bits. The information recovery is possible only when the original, unmarked data is available, offering more flexibility and security to the technology providers. The modification in this scheme could be brought by making the information dependent on the frequency characteristics of the host image (signal dependent information hiding) to make it more robust. It is expected from this work that it will make the data hiding technology so general and easier to implement with improved performances due to its fast response time and considerable transparency.

6. References

[1] Special issue on digital watermarking, *IEEE signal processing.*Mag.vol17, no.5, Sep.2000

[2] W. Zeng, "Digital watermarking and data hiding: technologies and applications," in Proc. Int. Conf. Inf. Syst., *Anal. Synth.*, vol. 3, 1998, pp. 223–229.

[3] Zhicheng Ni, Yun-Qing Shi, Nirwan Ansari, and Wei Su, “Reversible Data Hiding”, *IEEE transactions on circuits and systems for video technology*, vol. 16, no.3,March 2006.

[4] S.A.M. Gilani and A.N. Skodras, “Watermarking by multiresolution hamadard transform” (unpublished)

[5] D. kundur, “Multiresolution digital watermarking: Algorithms and implications for multimedia signals”, *PhD Thesis, university of Toronto*, 1999

[6] www.ece.utexas.edu/bevans/courses/ee381k/projects/fall98 /fu/report.pdf

[7] J. Cox, J. Kilian, T. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[8] A. Z. Tirkel, C. F. Osborne, and R. G. Van Schyndel, “Image watermarking-a spread spectrum application,” in Proc. *IEEE 4th Int. Symp. Spread Spectrum Techn. Applicat*., vol. 2, Sep. 1996, pp. 785–789.

[9] J. Irvine and D. Harle, Data Communications and Networks: An Engineering Approach. *New York: Wiley*, 2002

[10] F. Perez-Gonzlez and F. Balado, “Quantized projection data hiding,” in Proc. *IEEE Int. Conf. Image Process*., vol. 2, Sep. 2002, pp. 889–892.

[11] C. W. Honsinger, P. Jones, M. Rabbani, and J. C. Stoffel, “Lossless Recovery of an Original Image Containing Embedded Data,” *U.S. Patent 6 278 791 B1*, Aug. 21, 2001.

[12] B. Macq and F. Deweyand, “Trusted headers for medical images,” presented at the DFG VIII-D II Watermarking Workshop, *Erlangen, Germany*, Oct. 1999.

[13] M. Goljan, J. Fridrich, and R. Du, “Distortion-free data embedding,” in Proc. 4th Inf. Hiding Workshop, *Pittsburgh, PA*, Apr. 2001, pp. 27–41.