2005 IEEE International
Symposium on Signal Processing
and Information Technology

Binary Document Image Watermarking for Secure Authentication Using Perceptual Modeling

Niladri B. Puhan, Anthony T. S. Ho

Center for Information Security
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore, 639798
Email: niladri@pmail.ntu.edu.sg

Abstract - *In this paper, we propose a new perception based watermarking algorithm in binary document images for secure authentication purpose. Binary image watermarking with pixel flipping approach is a challenging problem, because flipping the black and white pixels in such simple images can bring noticeable visual distortion. A novel perceptual based model was proposed towards digital watermarking of binary images in [9]. The model estimates the distortion resulting from flipping of a pixel by finding the curvature-weighted distance difference (CWDD) measure between original and watermarked contour segments. In this paper, the reversible property of the CWDD measure is used towards designing a new authentication watermarking algorithm so that the possibility of any undetected modification to the watermarked image is removed. This algorithm embeds an authentication signature computed from the original image into itself after identifying an ordered set of low-distortion pixels. The same ordered set of pixels are correctly found in both the embedder and blind detector through the design of necessary conditions. The ability of the proposed authentication algorithm to detect any modification in the watermarked image is equivalent to the security of cryptographic authentication. The parity attack found in the previous block-wise data hiding methods in binary images is not possible in the proposed algorithm due to pixel-wise embedding of the authentication signature. Simulation results show the imperceptibility of the watermarking process and successful detection of content modifications.*

Keywords – Authentication, Binary image watermarking, Security, Perceptual model

I. INTRODUCTION

Digital watermarking is the art of protecting the multimedia data by inserting the proprietary mark which may be easily retrieved by the owner of the data to verify about its ownership or authenticity. A variety of digital watermarking methods have been developed for such purposes [1, 2]. For certain applications, watermarks for checking the authenticity of the multimedia data should be fragile because any corruption to watermarked data easily destroy the watermark and so the detection algorithm will be able to verify the integrity of the data being tested. Provable security of digital media can be guaranteed through the use of cryptographic signatures as the fragile watermark. Cryptographic signature has been well studied in cryptography and algorithms such as DSA, RSA and MD5 are extensively used in various authentication applications [3]. In authentication watermarking, the advantage of having the cryptographic signature hidden inside the digital data rather than appended to it is obvious. Lossless format conversion of the watermarked data does not render it inauthentic though the representation of the data is changed. Another advantage is that if the authentication information is localized, it is then possible to achieve the capability to localize the modifications after tampering by a hostile attacker. The present work involves watermarking in binary document images that could potentially include digitized versions of text, circuit diagrams, signature, driver licenses, financial and legal documents. By using different imaging software tools, the reproduction, distribution and editing of such images become easier. As such the ownership protection, authentication and annotation of binary images as well as tamper proofing have become necessary and important. Several watermarking methods in binary images have been proposed in literature.

Low *et al* [4, 5] proposed robust data hiding methods in formatted document images based on imperceptible line and word shifting. Their methods were applied to hide information in text images for bulk electronic publications. The line shifting method has low data hiding capacity as compared to the word shifting method but the hidden data is more robust to photocopying, scanning and printing process. Brassil *et al* proposed a scheme in [6], where the height of the bounding box enclosing a group of words could be used to hide data. This method has a better data hiding capacity than line and word shifting methods. Wu *et al* [7] hid data in a binary image using a hierarchical model in which human perception was taken into consideration. Distortion that occurred due to flipping of a pixel was measured by considering the change in smoothness and connectivity of a 3x3 window centered at the pixel. A single data bit is embedded in each block by modifying its total number of black pixels to be either odd or even. Shuffling was used to equalize the uneven embedding capacity over the image. Koch and Zhao [8] proposed a data hiding algorithm in which a data bit ‘1’ is embedded if the percentage of white pixels was greater than a given threshold, and a data bit ‘0’ is embedded if the percentage of white pixels was less than another given threshold. A sequence of contiguous or distributed blocks was modified by flipping the pixels until a certain threshold was reached. This algorithm was not robust

0-7803-9314-7/05/$20.00©2005 IEEE

to attacks and the hiding capacity was low. Our present work addresses the issue of secure authentication watermarking for binary images in electronic form in conjunction with cryptography techniques. Incorporating cryptography makes it possible to design a provably secure authentication watermarking scheme. Watermarking schemes for secure authentication applications need high capacity. Hiding considerable amount of data in binary images is a difficult problem, because minor modification can be relatively perceptible since the pixels are either black or white. A perceptual based *CWDD* model was proposed in [9] to minimize the visual distortion in the watermarked images. Through subjective experiments the *CWDD* model was found to be highly correlated with the human perception in estimating distortion due to the flipping of a pixel. By using this model it was possible to select suitable contour pixels for watermark embedding in binary images. The paper is organized as follows: in Section II, the motivation for proposing a new authentication watermarking algorithm in binary images is discussed. The proposed authentication watermarking algorithm is described in Section III for achieving provable security against any content modification. Results and discussion are presented in Section IV and finally some conclusions are given in Section V.

II. MOTIVATION

In this section, the motivation in proposing a new algorithm for authentication watermarking is discussed. Various authentication watermarking methods have been proposed for grayscale images in [10, 11]. However, there are only a limited number of authentication watermarking methods available for binary images. In a typical cryptography-based authentication watermarking method, an authentication signature (either message authentication code or digital signature) is computed from the whole image and embedded into the image itself. However, the very process of embedding a watermark alters the image, causing the subsequent authentication test to fail. To prevent this, it is usually necessary to partition the image into two parts, one of which is to be authenticated and the other to be altered to accommodate a watermark. A simple example is to partition an image such that the least-significant-bit (LSB) plane holds the authentication signature computed from the remaining bits of the image. However this idea does not work directly for binary images, because each pixel has only one bit. By modifying any pixel to embed a watermark would affect the signature of the image and the authentication test would fail. The challenging problem is how to divide the binary image into two parts such that the above idea of embedding the authentication signature can be applied. Cryptography-based authentication watermarking schemes were proposed for binary images in [12, 13]. Kim *et al* modified few bits in a binary image for embedding the authentication signature and the positions of those bits were known in both embedding and detection processes [12]. However this method of simple partitioning the binary image results in poor visual quality. In [13], Kim *et al* shuffled the binary image and then partitioned the shuffled image into two equal parts. Authentication signature was computed from one part and then embedded into the other part using the block-wise data hiding technique developed in [7]. A block in the second part of the image was embedded one bit of the authentication signature by modifying its total number of black pixels to be either odd or even. In this method, the first part is provably secure because the probability of undetected modification in this part is only 2^{-n} where n is the length of the authentication signature. However the second part of the image which carries the signature is prone to a 'parity attack'. The parity attack arises because the signature is embedded in the second part by considering the parity of the blocks, the number of black pixels. If two pixels that belong to the same block in the second part of the image change their values, the parity of this block may not change and so this modification will pass undetected. In Fig. 1 (a)–(d), the possibility of a parity attack to the watermarked image generated using the block-wise data hiding technique is illustrated. In the same paper, the proposed algorithm was modified to minimize the possibility of a parity attack. Thus each block in the second part of the image would have different probabilities of suffering due to parity attack and without being detected. As such this method is not provably secure against any type of modification to the watermarked image. In the next section, we propose a new authentication watermarking algorithm such that any modification in the watermarked image can be detected with high probability. The proposed algorithm achieves provable security because, (1) the original image can be partitioned into two parts by using the *CWDD* model and (2) the pixel-wise embedding of the authentication signature removes any possibility of parity attack.

III. PROPOSED AUTHENTICATION WATERMARKING ALGORITHM

In this section, we shall propose a new algorithm for authentication watermarking in binary images. After flipping a contour pixel, the amount of visible distortion can be estimated by the change in the contour segment that passes through the pixel. To calculate the distortion score for a contour pixel to be flipped, the 5-pixel length 'original contour segment' passing through this pixel is extracted by the contour tracing step [14]. Similarly after flipping the pixel, the 'watermarked contour segment' is also extracted. According to Eq. (6) in [9], the computation of *CWDD* measure takes the original and watermarked contour segments into account. The *CWDD* measure of a pixel remains the same before and after its flipping, because it is computed by considering the change occurred during the flipping process. This reversible property is particularly useful in identifying an ordered set of pixels in the original image. During watermarking each such pixel can carry one bit of the authentication signature computed from the remaining pixels in the image. For blind detection, these pixels should be detected in the same order as the embedder during the watermarking process. However, direct application of the reversible property of the distortion measure is not sufficient alone for this purpose. Certain necessary conditions are designed for the correct detection of the ordered set of pixels in the proposed algorithm. We define each such pixel as the reversible pixel in an image. If a set of reversible pixels are found, the original image can be divided into two parts as necessary for correctly embedding the authentication signature. Secondly, the pixel-wise embedding of the authentication signature in the reversible pixels shall remove the possibility of parity attack. The following steps explain the proposed authentication watermarking algorithm in binary images.

Embedding:

1. The *CWDD* measure for each contour pixel in the original image is computed. The contour pixels having the distortion score below a threshold *T* are defined as the suitable pixel and rest of the pixels in the image are defined as non-suitable. The suitable pixels after flipping bring less visible distortion to the watermarked image and in the next step only these pixels are examined for watermarking.
2. Among the suitable pixels, *N* numbers of reversible pixels that satisfy a set of conditions are searched in a sequential scanning order starting from left to right and top to bottom of the original image. The conditions for a suitable pixel to be defined as the reversible pixel are given below:
   a. In an *MxM* pixel window centered on the current suitable pixel, there should not be any reversible pixel already found in the image.
   b. If after flipping, it does not change the suitability status of any neighboring pixel which comes before it in the scanning order. A change in suitability status means that a suitable pixel is changed to a non-suitable pixel and vice-versa after the flipping process. The neighborhood considered is a *5x5* pixel window centered on the current suitable pixel.

   Condition (a) is necessary because flipping of the current suitable pixel may cause a change in the status of the reversible pixels near to it. Condition (b) is necessary because any change in the suitable status of the neighbor pixels coming before it in scanning order may lead to an error in the blind detection. However, any change in the suitability status of the pixels subsequently in the scanning order does not cause any error because of condition (a). To verify any such change in their suitability status, the *CWDD* measure for the neighbor pixels is computed after flipping the center pixel. In Fig. 2, the condition 2 (b) is shown as an example.
3. After *N* reversible pixels are found, all pixels in the original image are divided into two disjoint subsets. The reversible pixels form the reversible subset S_R and remaining pixels in the original image belong to the message subset S_M .
4. The authentication signature is computed from the pixels in S_M using the secret key and embedded into the pixels of the reversible subset. The embedding is performed pixel-wise; so each reversible pixel in S_R holds one bit of the authentication signature and the reversible pixel value is set equal to the signature bit it holds.
5. Set union operation of the embedded reversible subset $S_R{}^m$ and the message subset S_M generates the watermarked image.

Detection:

6. Similar to the steps 1, 2 and 3 in the embedding process, *N* numbers of reversible pixels are searched in the test image at the blind detector in sequential scanning order and all pixels in this image are divided into two disjoint subsets. The reversible pixels form the reversible subset $S_R{}'$ and remaining pixels in the test image belong to the message subset $S_M{}'$.
7. The *N*-bit authentication signature is computed from the pixels in $S_M{}'$ using the secret key and compared with the *N* reversible pixel values of $S_R{}'$. If each signature bit matches with the corresponding reversible pixel value, then the image under question is authentic. Otherwise this image has been modified after the watermarking process.

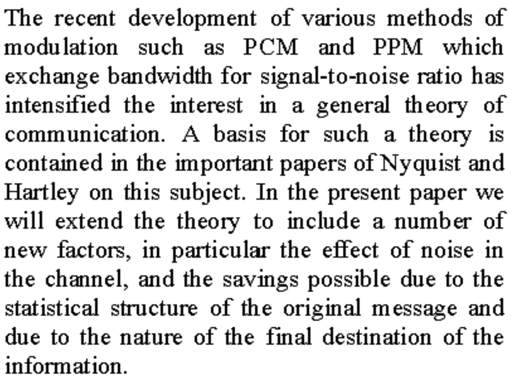
IV. RESULTS AND DISCUSSION

In this section, we present the simulation results by implementing the authentication watermarking algorithm proposed in the previous section. The authentication signature to be used in this algorithm is the Hashed Message Authentication Code (HMAC). The HMAC is found by computing the one way hash function of the data string that is a concatenation of the pixel set and secret key. In our method, provable security against any modification is obtained by using the cryptographic hash function. In the implementation, we have used MD5 [15] hash function to compute the HMAC and the output 128-bit HMAC is used as the authentication signature. The original image of size 320x432 pixels in Fig. 1(a) is used to demonstrate the effectiveness of the algorithm against content modification. Here *N* is equal to 128 and a total of 128 reversible pixels are searched in the original image by following the sequential scanning order. The threshold parameter *T* for choosing the suitable pixel by the *CWDD* measure is 0.5 and the parameter *M* is chosen to be 11 for keeping the reversible pixels separated by a distance. Low value of parameter *T* brings less visual distortion in watermarked image and parameter *M* is chosen to be 11 to satisfy condition 2(a) for correct blind detection. The choice of higher *M* value reduces visual interference among the reversible (flipping) pixels being separated by a larger distance. Thus visual quality of the watermarked image is less affected. However, the number of available reversible pixels is reduced with the increase in *M* value. Fig. 3 shows the watermarked image after embedding the 128-bit HMAC which is perceptually similar to the original image. In Fig. 4, 128 reversible pixel positions in the original and watermarked image are shown to be identical. Since the position map of the reversible pixels in both the images is identical, correct blind detection is possible after the watermarking process. In our simulation, a maximum of 584 reversible pixels are found in both the images using the chosen parameters. We perform multiple modifications such as deletion, insertion and substitution of characters in the watermarked image; (1) the only word *'information.'* in last line is deleted, (2) the word 'theory' is inserted in the last line, and (3) the word 'for' in line-5 is substituted by the word 'to'. The resulting attacked image is shown in Fig. 5 (a). At the detector side the attacked image fails in the authentication test. In Fig. 5(b), difference between the HMAC computed from the message subset and the reversible subset illustrate the failure of the attacked image to pass the authenticity test.

The performance of the proposed algorithm can be compared with the previous three methods that also address the issue of authentication watermarking in binary images [7, 12, 13]. The method suggested by Wu *et al* in [7] detects the modification by using a block-based data hiding technique after shuffling the binary image. Though this method can detect tampering, it suffers from the parity attack and here security level is less as comparable to the authentication methods that use the cryptographic signatures. In [12], Kim *et al* performs pixel-wise embedding of the cryptographic signature, so their method is not vulnerable to parity attack and can detect any modification with high probability. However the visual quality of the watermarked image becomes poor because necessary perceptual modeling is not performed. In [13], Kim *et al* proposes a method to detect any modification in binary images by embedding the cryptographic signature block-wise. As discussed earlier this is also vulnerable to parity attack but the visual quality is not reduced after watermarking. The security of the proposed authentication algorithm can be directly related to a secure cryptographic element such as a hash function. The probability of any undetected modification in the watermarked image is only 2^{-n} where *n* is the length of the authentication signature being used. If the attacker wants to modify the message subset such that the authentication signature remains same, the chances of obtaining such a collision are removed by using a secure cryptographic key of length 128 bits or more. Further, if the attacker alters any signature bearing reversible pixel, then the computed signature from the message subset will not match with the pixels in the reversible subset. Use of the *CWDD* measure ensures the visual quality of the watermarked image remains perceptually similar to the original image. The possibility of parity attack is not present here because each bit of authentication signature is carried by a reversible pixel instead of a block. To summarize, the ability of the proposed authentication algorithm for detecting any type of content modification in the watermarked image is equivalent to the security of cryptographic authentication without being susceptible to any attacks.

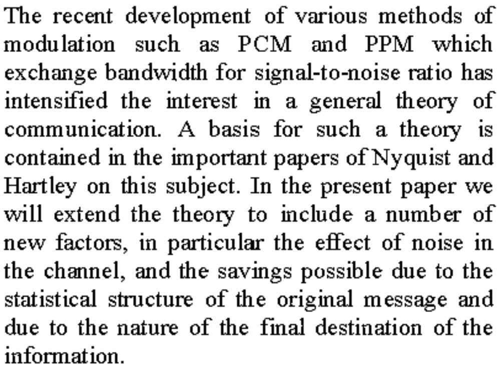
V. Conclusion

In this paper, we proposed a new authentication watermarking algorithm for binary images. The proposed algorithm can detect any kind of modification to the watermarked image with probability equivalent to that of the cryptographic authentication. For this purpose, an ordered set of low-distortion reversible pixels are selected for pixel-wise embedding of the authentication signature. This selection of the reversible pixels is performed by designing the necessary conditions and using the reversible property of *CWDD* measure. The proposed algorithm does not suffer from any parity attack like the other block-wise data hiding methods in binary images. The application of the proposed algorithm for binary images can be used in secure FAX transmission. After a transmission is performed, the sending FAX machine embeds the watermark using its own secret-key. The receiver Fax machine can verify the received document whether it has not been modified after the transmission. Another application could be the legal usage of binary documents. If documents are stored in a database, the user can verify their authenticity by using the appropriate key.



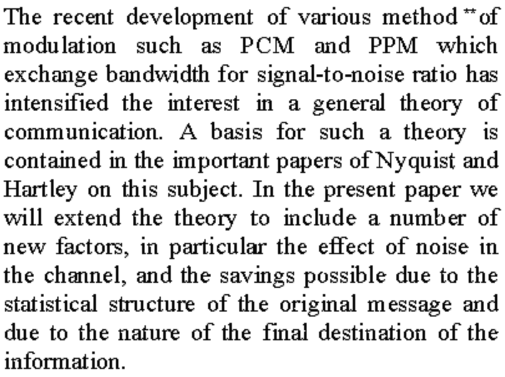
The recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist and Hartley on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

(a)



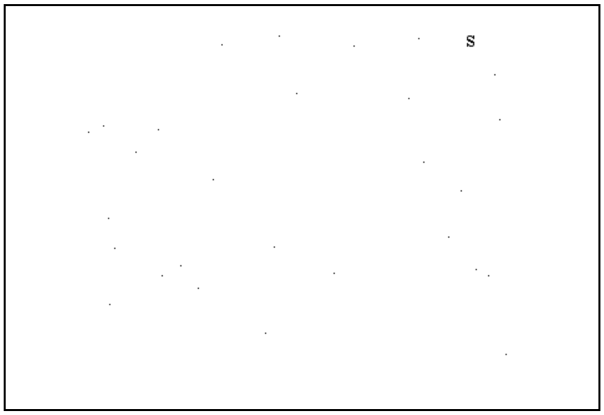
The recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist and Hartley on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

(b)



The recent development of various method **of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist and Hartley on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

(c)

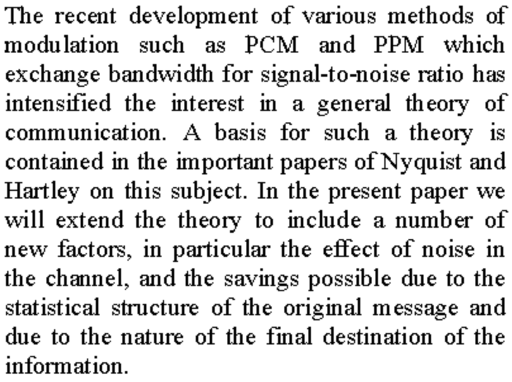


(d)

Fig. 1. Example showing the parity attack (a) the original image of size 320x432 pixels (b) the watermarked image (c) the fake image is created by deleting a character 'S' at the ** marked position and with 28 extra pixels flipped in the watermarked image. Both images in (b) & (c) give the same detection output (d) image showing difference between the watermarked image and fake image.

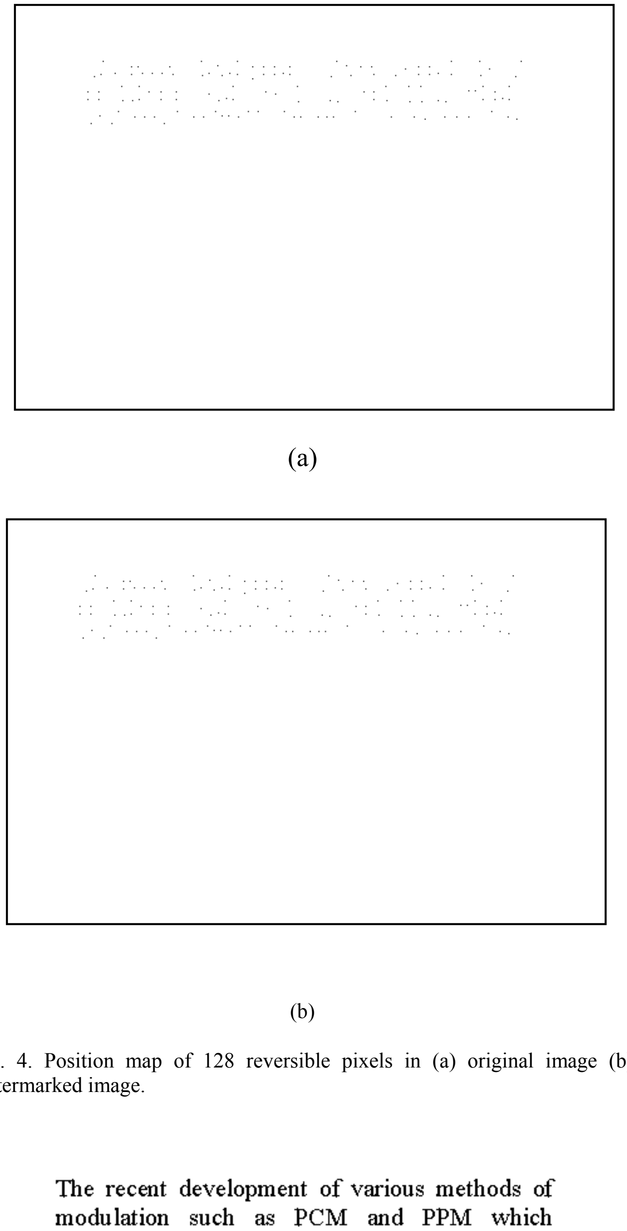
434	**435**	**436**	**437**	**438**
866	**867**	**868**	**869**	**870**
1298	**1299**	*1300*	1301	1302
1730	1731	1732	1733	1734
2162	2163	2164	2165	2166

Fig. 2. Condition 2(b) is illustrated as an example for the center pixel at 4th row and column in the original image. The scanning order of the center pixel and its *5x5* pixel neighborhood in the image are shown. Pixels in bold case are checked to detect any change of their suitability status after flipping the center pixel.



The recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist and Hartley on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

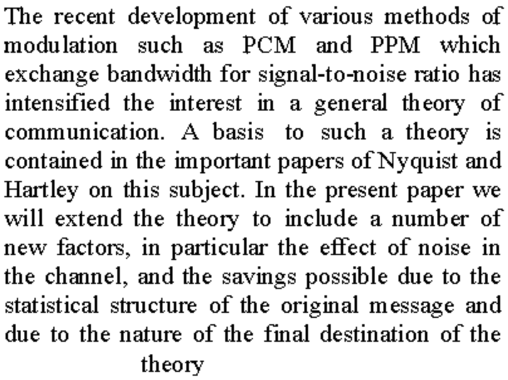
Fig. 3. Watermarked image with 128-bit HMAC embedded in the original image.



(a)

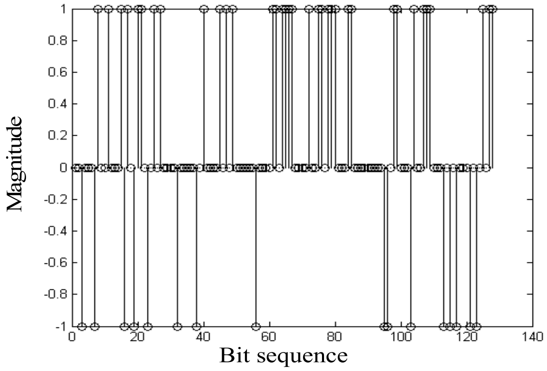
(b)

Fig. 4. Position map of 128 reversible pixels in (a) original image (b) watermarked image.



The recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis to such a theory is contained in the important papers of Nyquist and Hartley on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the theory

(a)



(b)

Fig. 5. (a) Attacked image, (b) difference between the HMAC and the reversible subset of the attacked image.

REFERENCES

[1] M. D. Swanson, M. Kobayashi, A. H. Tewfik, “Multimedia Data-Embedding and Watermarking Technologies, ” *Proc. of the IEEE*, vol. 86, no. 6, June 1998.

[2] I. J. Cox and M. L. Miller, “A Review of Watermarking and the Importance of Perceptual Modeling,” *Proc. SPIE*, vol. 3016, Feb. 1999.

[3] A. Menezes, P. van Oorchot, S. Vanstone, “Handbook of Applied Cryptography,” Boca Raton, FL: CRC, 1997.

[4] S. H. Low, N. F. Maxemchuk, and A. M. Lapone, “Document identification for copyright protection using centroid detection,” *IEEE Trans. on Communication,* vol. 46, no. 3, March 1998, pp. 372-383.

[5] S. H. Low, and N. F. Maxemchuk, “Performance comparison of two text marking methods,” *IEEE Journal on Selected Areas in Communications,* vol. 16, no. 4, May 1998.

[6] J. Brassil and L. O’Gorman, “Watermarking document images with bounding box expansion,” *Proc. 1st Int’l Workshop on Information Hiding,* Newton Institute, Cambridge, UK, May 1996, pp. 227-235.

[7] M. Wu, E. Tang, and B. Liu, “Data hiding in digital binary images,” *Proc. IEEE Int’l Conf. on Multimedia and Expo,* Jul 31-Aug 2, 2000, New York.

[8] E. Koch, J. Zhao, “Embedding robust labels into images for copyright protection, ” *Proc. International Congress on Intellectual Property Rights for Specialized Information, Knowledge & New Technologies,* Vienna, Aug. 1995.

[9] A.T.S. Ho, N. B. Puhan, P. Marziliano, A. Makur, Y. L. Guan, “Perception Based Binary Image Watermarking,” *IEEE International Symposium on Circuits and Systems* (ISCAS), Vancouver, Canada, 23-26 May (2004).

[10] P. Wong, “A Watermark for Image Integrity and Ownership Verification,” *Proc. IS&T PIC*, Portland, Oregon, 1998.

[11] P.W. Wong and N. Memon, “Secret and Public Key Image Watermarking Schemes for Image Authentication and Ownership Verification,” *IEEE Trans. Image Processing*, vol. 10, no. 10, October 2001.

[12] H. Y. Kim and A. Afif, “Secure Authentication Watermarking for Binary Images,” *Proc. Sibgraphi – Brazilian Symposium on Computer Graphics and Image Processing*, pp 199-206, 2003.

[13] H. Y. Kim and R. L. de Queiroz, “Alteration-Locating Authentication Watermarking for Binary Images,” *Proc. Int. Workshop on Digital Watermarking 2004,* (Seoul), LNCS-2939, 2004.

[14] Abeer George Ghuneim, “Tutorial on Contour Tracing,” available at *http://www.cs.mcgill.ca/~aghnei/index.html.*

[15] R. L. Rivest, “RFC 1321: The MD5 Message-Digest Algorithm,” *Internet Activities Board*, 1992.