# A Semi-Fragile Pinned Sine Transform Watermarking System For Content Authentication of Satellite Images

Anthony T. S. Ho, Xunzhan Zhu
School of Electrical and Electronic Engineering
Nanyang Technological University
Email: etsho@ntu.edu.sg

W. M. Woon
DataMark Technologies Pte Ltd,
100 Jurong East Street 21,
ST Electronics Jurong East Building, Singapore

***Abstract*— A novel semi-fragile watermarking scheme for the content authentication of satellite images using the pinned sine transform (PST) is presented in this paper. In the PST domain, the image field is first decomposed into two mutually orthogonal sub-fields, namely, the boundary field and the pinned field. The watermark is embedded into the pinned field of PST, which contains the texture information of the original image. This important property of the pinned field provides the scheme with special sensitivity to any texture alteration to the watermarked image. It is desirable to the authentication of satellite images, to which the texture characters are crucial for semantic understanding. The watermarking system can localize the portions of a watermarked image that have been tampered maliciously with high accuracy as well as approximately recover it, while maintaining a high degree of image integrity. The inter-block relationship introduced in the process of PST renders the watermarking scheme resistant to content cutting-and-pasting attacks. The watermark can still survive slight non-malicious manipulations, which is desirable in many practical applications in satellite remote sensing. Simulation results demonstrated that the probability of tamper detection of this authentication scheme is higher than $98\%$.**

## I. Introduction

With the tremendous development of the Internet, the online purchasing and distribution of satellite images can now be performed relatively easily. It follows that the content authentication or the integrity protection of digital satellite images is increasingly becoming more and more important. To counter this growing information security problem of illegal distribution and counterfeiting, a novel semi-fragile watermarking scheme for the content authentication of satellite images using the pinned sine transform (PST) [1], [2] is presented in this paper, which is based on our previous work [3].

The original image is decomposed into two mutually uncorrelated fields, namely, the boundary field and the pinned field. The texture information of the original image is contained in the pinned field, wherein the sine transform is equivalent to a fast Karhunen-Loeve transform (KLT). By exploiting this important property, we propose to embed a watermark signal into the sine transform domain of the pinned field for content authentication. As illustrated in this paper, the proposed watermarking scheme is especially sensitive to texture alterations of the host image. This provides significant advantage for authentication of biomedical images, which is strongly texture based. Moreover, although our scheme is block-wise, the watermarking of one block is closely related to all the blocks surrounding it, which renders our scheme robust to the "cutting and pasting" attacks. The next section presents a brief review of the PST. The watermark embedding and authentication methods are elaborated in Section III and Section IV, respectively. Section V gives the simulation results followed by the conclusion in Section VI.

## II. Pinned Sine Transform

Suppose a data vector

$$\mathbf{X} = [x_0 \ \ldots \ x_{n+1}]^T$$

is separated into a boundary response $\mathbf{X}^b$ defined by $x_0$ and $x_{n+1}$, and a residual sequence $\mathbf{X}' - \mathbf{X}^b$, where

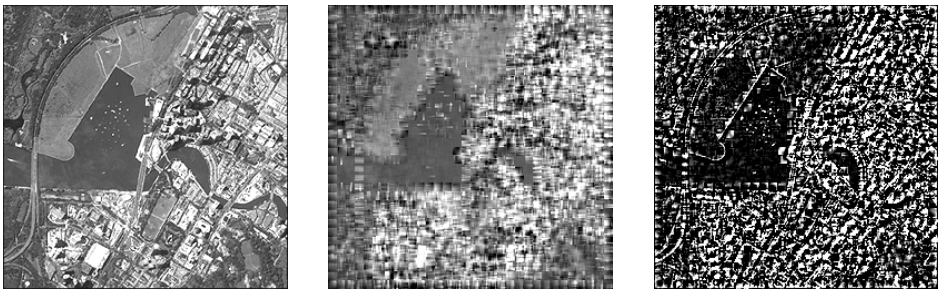
$$\mathbf{X}' = [x_1 \ldots \ x_n]^T.$$

In [4], Jain showed that if $\mathbf{X}$ was a first-order stationary Gauss-Markov sequence, the sequence $\mathbf{X}' - \mathbf{X}^b$ had the sine transform as its KLT.

Extending the above theory to the more general 2-D case, Meiri et. al. [1], [2] proposed the decomposition of an image field into two sub-fields, namely, the boundary field and a residual field. The boundary field depends only on the block boundaries and for the residual field, so-called the pinned field in [2], which vanishes at the boundaries, its KLT is the sine transform.

The semi-fragile watermarking adapts to a selective authentication scheme on the content of images. Our scheme aims at protecting the primary textures, such as edges, of the biomedical images. To this end, the watermark should not survive the authentication process if such textures are tampered or damaged. The results of the decomposition of the $512 \times 512$ *Singapore* image are shown in Figure 1. We find that the boundary field is only a blurred version of the original image, while the pinned field is a good characterization of edges, which fully reflects the texture information in the original image. Thus the watermark can be embedded into the pinned

0-7803-9050-4/05/$20.00 ©2005 IEEE.



(a) (b) (c)

Fig. 1. The dual-field decomposition of the *Singapore* image: (a) the original image, (b) the boundary field and (c) the pinned field.

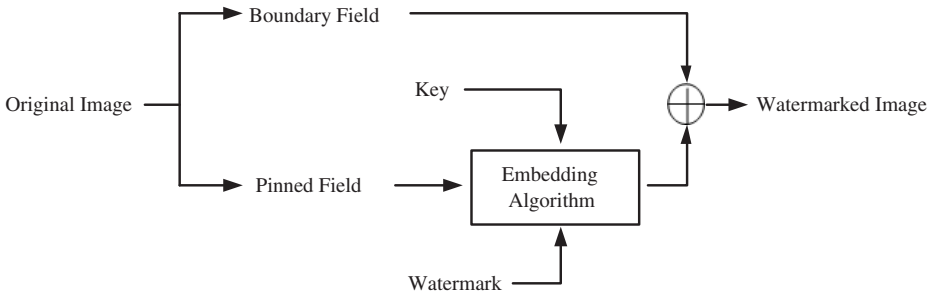



Fig. 2. Watermark embedding process.

field as an indicator of the authenticity of the watermarked image. Moreover, since most common image manipulations tend to preserve such primary features of images, this embedding method ensures that the watermark does not suffer significantly from such legitimate manipulations.

In PST, the image is divided into overlapped blocks, which introduces an inter-block relationship to the pinned sine transformed images. Therefore the watermarking of any particular block also depends on its location in the image instead of depending only on its own content. Thus, simple "cutting and pasting" counterfeiting attack can be exposed by this encoding scheme since the counterfeit of one block affects all the blocks around it.

## III. Watermark Embedding

The semi-fragile watermarking aims at authentication of the semantic content of images, i.e., they should protect the primary features of the image content, e.g., edges. The watermark should not survive if such features are damaged. As we have seen in the previous section, after the dual-component decomposition, information of prominent edges of the original image is contained in the pinned field. Thus we may embed the watermark into the pinned field as an indication of the authenticity of the watermarked image.

The watermark embedding process is described in Figure 2. The details are described as follows. The original image $\mathbf{X}$ is partitioned into overlapping blocks $\{\mathbf{X}_{m,n}\} \in \mathbf{X}$ of size $10 \times 10$, where $m$ and $n$ are the coordinate numbers of this block. Two neighboring blocks are overlapped by one column or row. For every block, the surrounding zone of a 2-pixel width is averaged to generate the initial boundaries and corners. These parameters are used to achieve the boundary field by interpolation and the pinned field is in turn obtained by subtracting the boundary field from central $8 \times 8$ part of the original block. After every block has been decomposed, it results in non-overlapping pinned field blocks and boundary field blocks, denoted as $\{\mathbf{X}^p_{m,n}\}$ and $\{\mathbf{X}^b_{m,n}\}$, respectively. Both the boundary field and the pinned field are of size of $8 \times 8$.

The watermarking process proceeds by conducting the sine transform to every $\mathbf{X}^p_{m,n}$ block and by embedding a pseudo-random binary sequence of length $L$ into each block, whose initial seed is contained in a secret key file. In the middle to high frequency bands of the sine transform coefficients, we select, according to the length of the watermark sequence, $L$ coefficients for watermarking modulation. More specifically, the watermarking process is defined as follow:

$$y^p = \begin{cases} x^p & (w = 1 \wedge x^p > T) \vee (w = 0 \wedge x^p < -T) \\ \alpha_1 & w = 1 \wedge x^p \leq T \\ \alpha_2 & w = 0 \wedge x^p \geq -T, \end{cases} \tag{1}$$

where $x^p$ and $y^p$ are the coefficients before and after water-

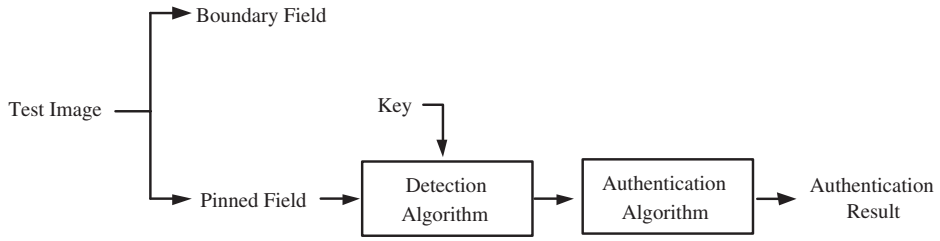



Fig. 3. Watermark detection and image authentication process.

marking, respectively. $T$ is a sufficiently large positive threshold value, which is set to achieve the best tradeoff between the perceptual quality and robustness. $\alpha_1$ and $\alpha_2$ are floating point values with $\alpha_1 \in [T/2, T]$ and $\alpha_2 \in [-T, -T/2]$.

The watermarked coefficients are then inverse sine transformed and a watermarked image is obtained by adding the boundary field to the watermarked pinned field.

## IV. Watermark Detection and Image Authentication

The detection of watermark is performed as follows. The detection system receives as input a watermarked and possibly tampered image $\hat{\mathbf{Y}}$. After a similar block-wise decomposition as in the watermark embedding, we obtain the pinned field blocks $\{\hat{\mathbf{Y}}^p_{m,n}\}$. A Sine transform is performed on these blocks. The watermarked coefficients are then located and checked based on the following conditions: if $\hat{y}^p > 0$, we decide the watermark bit as "1"; otherwise, we decide it as "0".

After collecting all the watermark bits in one block, we obtain the retrieved and possibly corrupted watermark. The original watermark is also generated using the initial seed in the key file. The watermark bits are compared via the normalized cross correlation function:

$$\rho = \frac{\sum_{l=0}^{L} \hat{w}_{m,n}[l] w_{m,n}[l]}{\left[\sum_{l=0}^{L} (\hat{w}_{m,n}[l])^2\right]^{1/2} \left[\sum_{l=0}^{L} (w_{m,n}[l])^2\right]^{1/2}}, \quad (2)$$

where $w_{m,n}$ is the watermark signal and $\rho \in [-1, 1]$. The integrity of the block $\hat{\mathbf{Y}}_{m,n}$ is evaluated according to the value of $\rho$. If no tampering ever occurred to this block, $\rho \to 1$; on the other hand, $\rho$ will decrease due to different tampering of $\hat{\mathbf{Y}}_{m,n}$. If the content of the block has been changed, i.e. the block has been replaced, due to properties of the normalized cross correlation function, $\rho$ will be extremely low.

Assume $\gamma$ is a properly set threshold, the block is considered to be maliciously tampered if $\rho < \gamma$. The threshold is determined mathematically or experimentally so as to maximize the probability of detection subject to a given probability of false alarm. In our current simulations, $\gamma$ is experimentally set to tolerate unavoidable non-malicious modifications in some practical applications, such as JPEG compression and noise addition, while maintain the sensitivity of the authentication process to malicious modification on the content of the watermarked images. When tampered portions are detected, these portions can be recovered using the similar method as that in [5].

## V. Simulation Results

We use two $512 \times 512$ gray-scale satellite images *Singapore* and *Pyramids* as shown in Figures 4(a) and 5(a) to test our authentication algorithm. Figures 4(b) and 5(b) display the watermarked images. We can see that the watermarked images look identical to the original images, with PSNR values of approximately 40 dB and 38 dB, respectively. We modified the content of the watermarked images as follows: in Figure 4(c), the bridge was removed and in Figure 5(c), some texture of the image was modified. Furthermore, to illustrate the insensitivity of our algorithm to non-malicious modifications, after content modification, the images were saved as JPEG files with a QF of $90\%$. As illustrated in Figures 4(d) and 5(d), the maliciously modified areas were accurately detected and identified, while the normal image processing manipulations have little affect on the authentication process. The restoration results are also shown in Figures 4(e) and 5(e).

## VI. Conclusion

In this paper, we discuss a novel semi-fragile watermarking using the pinned sine transform for the authentication of satellite images. The watermark is embedded into the pinned field, which contains the texture information of the original image. This important property of the pinned field provides the scheme with special sensitivity to any texture alteration to the biomedical images. The watermarking system can localize the portions of a watermarked image that have been tampered maliciously with high accuracy. The effectiveness of the new method has been demonstrated by experimental results.

## References

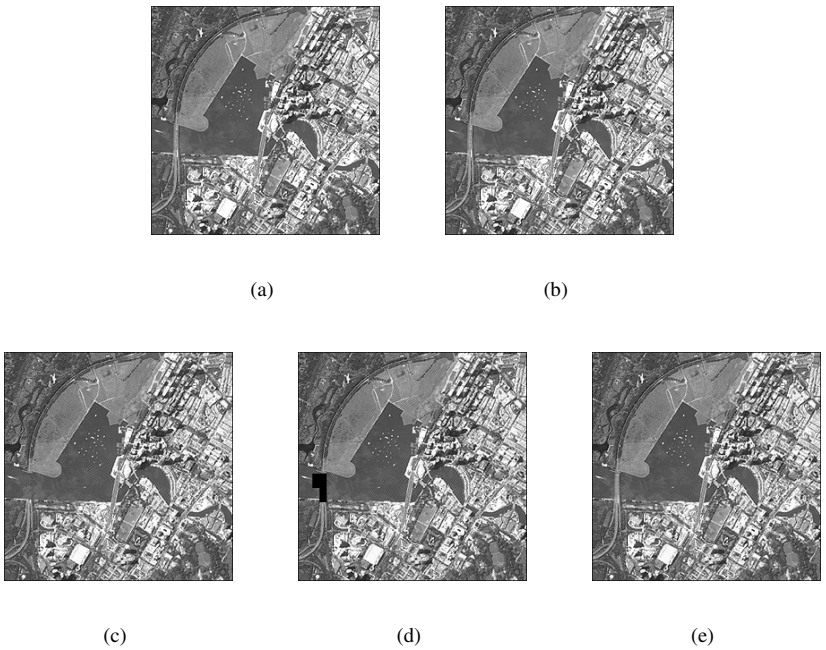
[1] A. Z. Meiri, "The pinned Karhunen-Loeve transform of a two dimensional Gauss-Markov field," in *Proc. SPIE Conf. Image Processing*, San Diego, CA, 1976.

[2] A. Z. Meiri and E. Yudilevich, "A pinned sine transform image coder," *IEEE Trans. Commun.*, vol. COM-29, pp. 1728–1753, Dec. 1981.

[3] A. T. S. Ho, X. Zhu, and Y. L. Guan, "Image content authentication using pinned sine transform," *EURASIP Journal on Applied Signal Processing (JASP), Special Issue on Multimedia Security and Rights Management*, vol. 2004, no. 14, pp. 2174–2184, Oct. 2004.

[4] A. Jain, "Some new techniques in image processing," in *Image Science Mathematics*, O. Wilde and E. Barrett, Eds. California: Western Period., 1976.

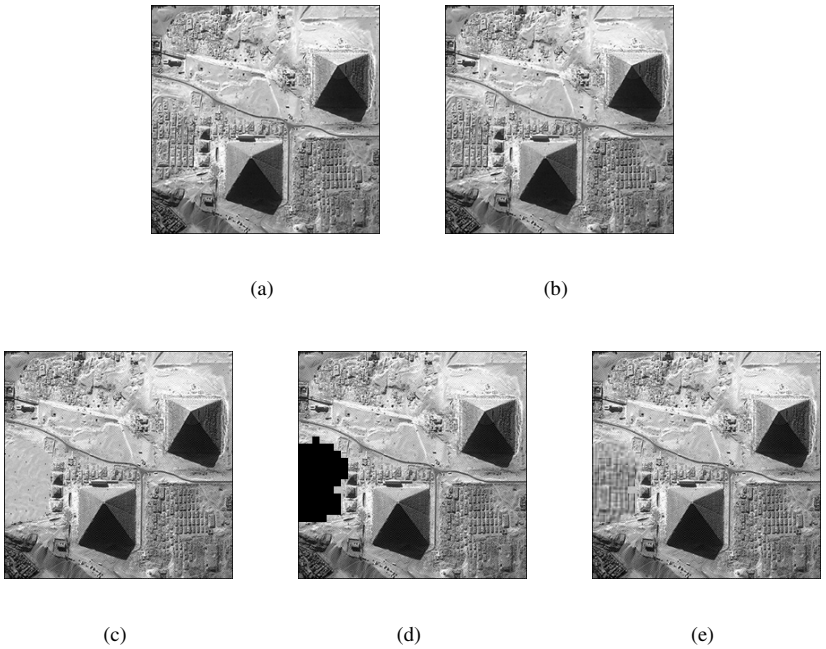
[5] J. Fridrich and M. Goljan, "Images with self-correcting capabilities," in *IEEE International Conf. on Image Processing*, Kobe, Japan, Oct. 1999, pp. 792–796.



(a) (b)

(c) (d) (e)

Fig. 4. (a) The original image *Singapore*; (b) the watermarked image; (c) the tampered image (the bridge was removed); (d) the authentication result and (e) the restoration result.



(a) (b)

(c) (d) (e)

Fig. 5. (a) The original image *Pyramids*; (b) the watermarked image; (c) the tampered image (the texture was modified); (d) the authentication result and (e) the restoration result.