

Semi-fragile Watermarking and Authentication for Law Enforcement Applications

Anthony T.S. Ho

Department of Computing, School of Electronics and Physical Sciences, University of Surrey,
UK

Email: a.ho@surrey.ac.uk

Keywords: Digital watermarking, authentication, pinned sine transform, crime scene analysis.

Abstract

In this paper, a semi-fragile watermarking method for authentication of law enforcement images such as digital images captured at crime scenes and traffic enforcement situations is proposed using the pinned sine transform (PST). The watermarking system can localize the portions of image that have been tampered maliciously, with high accuracy. In particular, the watermarking scheme is very sensitive to any texture alteration in the watermarked images, which is crucial for crime scene image authentication. Simulation results are presented to demonstrate the effectiveness of the proposed method and its possible applications in the field of crime scene analysis. The proposed method of watermarking authentication could potentially prove useful when digital photographs are presented as evidence in the court of law.

1. Introduction

Crime scene photography plays significant role on visually capturing a crime scene for possible later use of the photographs as evidence in the court of law. The role of a scene of crime officer (SoCOs) is to capture, as much as possible, the left-over evidence at the crime scene, by taking photographs and collecting any exhibits found. After the collection of evidence, there is no other way of examining the crime scene as a whole, apart from looking back at the collected exhibits and photographs taken. Crime scene photography comprises three different kinds of photographs: photographers first try to capture a “general” shot of the whole scene, then the photographer moves in towards particular objects and captures “mid-range” shots of the scene, and finally “close up” shots are

taken of a particular piece of evidence. Moveable exhibits usually are taken back to the studio to be photographed from several angles [1]. Therefore, it is very important that all photographic evidence remains unchanged and authentic.

A semi-fragile watermarking is a potential solution to the image content authentication problem which seeks to verify that the content of the multimedia has not been modified by any of a predefined set of illegitimate distortions, while allowing modification by legitimate distortions [2].

In this paper, we propose a novel semi-fragile watermarking scheme for crime scene image authentication using the pinned sine transform (PST), 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 crime scene images, whose analysis 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 the proposed watermarking and authentication method for crime scene analysis. Section 3 gives the simulation results which illustrate proposed possible applications of our method, followed by the conclusion in Section 4.

2. The proposed method

In this section, we describe our proposed watermarking method for crime scene image authentication. Some of the photographs used here are taken by SoCO photographers at the Metropolitan Police Scientific Support College training centre in Hendon. A special mock crime scene was set up, and an investigation was carried out, as in a real crime scene, where the SoCOs had to collect all exhibits and photograph the crime scene. A total number of 64 photographs were taken to capture the crime scene, containing general, mid-range, and close-up shots. Selected photographs were then watermarked and used for the purpose of our experiments. Section 2.1 gives a short review of the pinned sine transform. The watermarking and authentication processes are presented in Sections 2.2 and 2.3, respectively.

2.1 Pinned sine transform

Suppose a data vector

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

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

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

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. [5, 6] 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 [6], which vanishes at the boundaries, its KLT is the sine transform. After decomposition, 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 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.

2.2 Watermark embedding

The watermark embedding process is described as follows. The original image \mathbf{X} is partitioned into overlapping blocks $\{\mathbf{X}_{m,n}\} \in \mathbf{X}$ of size 10×10 , where m and n are the coordinate numbers of this block. Two neighbouring 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 8×8 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 $\{\mathbf{X}_{m,n}^p\}$ and $\{\mathbf{X}_{m,n}^b\}$, respectively. Both the boundary field and the pinned field are of size of 8×8 .

The watermarking process proceeds by conducting the sine transform to every $\{\mathbf{X}_{m,n}^p\}$ 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:

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

where x^p and y^p are the coefficients before and after watermarking, respectively. T is a sufficiently large positive threshold value, which is set to achieve the best trade-off between the perceptual quality and robustness. α_1 and α_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.

2.3 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 $\{\hat{\mathbf{Y}}_{m,n}^p\}$. 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 ρ . If no tampering ever occurred to this block, $\rho \rightarrow 1$; on the other hand, ρ 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, ρ will be extremely low.

Assume γ 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 investigation, γ is experimentally set to tolerate unavoidable non-malicious modifications in some practical applications, such as JPEG compression and noise addition, while maintaining the sensitivity of the authentication process to malicious modification on the content of the watermarked images. When tampered portion is detected, the portion can be recovered using the method of projection [7].

3. Proposed Applications

In this section, we show some experimental results which can be used to illustrate the possible applications of our method. Figure 1(a) shows a typical mid-range crime scene image, and its watermarked version is shown in Figure 1(b). We can see that the watermarked

images look identical to the original images, with PSNR of 35 dB. We modified the content of the watermarked images by cutting and pasting blocks in the same watermarked images as displayed in Figure 1(c): the gun on the table was removed maliciously. The authentication and restoration results are presented in Figure 1(d) and Figure 1(e), respectively. We can find that the modified areas were accurately detected and identified, and the approximately recovered image is visually acceptable. Similarly, Figure 2 shows an example of fingerprint authentication, and Figure 3 is an example of traffic image authentication for traffic enforcement. The Probability of tamper detection and the quality of the recovered portions are also tabulated in Table 1.

Table 1: Probability of tamper detection and quality of recovered portions

Image	P_{TD} (%)	PSNR
Figure1	98	28.99
Figure2	99	23.54
Figure3	99	27.95

The probability of tamper detection is defined as:

$$P_{TD} = \frac{NUM_{detected}}{NUM_{modified}}, \quad (3)$$

Where $NUM_{modified}$ is the number of actually modified blocks and $NUM_{detected}$ is the number of correctly detected blocks. The quality of recovered portions is evaluated by peak signal to noise ratio (PSNR) compared to the original images, which is defined as:

$$PSNR = 20 \log_{10} \left(\frac{255 \times N^2}{\sum [X_r(i,j) - X(i,j)]^2} \right), \quad (4)$$

Where $X_r(i,j)$ and $X(i,j)$ are the pixels of the recovered images and those of the original image, respectively.

4. Conclusion

In this paper, we discussed a novel semi-fragile watermarking algorithm using the pinned sine transform for the authentication of crime scene images. The watermark was embedded into the pinned field, which contained 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 crime scene images. The

watermarking system can localize the portions of a watermarked image that have been tampered maliciously with high accuracy of probability of tampered detection to be approximately 99%, as well as providing the capability for self-correction of the tampered regions to PSNR of 29dB of restored image quality as indicated in Table 1. This method of authentication could potentially solve, or at least add more credibility, to the problem of using digital photography as evidence in court.

For future work, the proposed watermarking scheme can be adapted to link the different photograph shots (general, mid-range and close up) through embedding one image to another. This would provide a common secure linkage, thus ensuring that the images have not been tampered with throughout the flow process from point of capture to database archival. Another alternative would be to create photograph identification numbers as hyperlinks of the set of images and embedding them as watermarks in the images.

5. References

- [1] B. Vrusias, M. Tariq, C. Handy, S. Bird, "Forensic Photography", Technical Report, University of Surrey, Computing Dept., 2001.
- [2] I. J. Cox, M. L. Miller and J. A. Bloom. "Digital Watermarking", San Francisco, Calif, USA: Morgan Kaufman Publishers, (2001).
- [2] A. T. S. Ho, X. Zhu and Y. L. Guan. "Image content authentication using pinned sine transform". *EURASIP J. Appl. Signal Process, Spec. Iss. Multimedia Secur. & Rights Manage*, vol. 2004, pp. 2174 – 2184, (2004).
- [3] A. K. Jain. "Some new techniques in image processing". *Image Science Mathematics*, O. Wilde and E. Barrett, Eds. California: Western Period (1976).
- [4] A. Z. Meiri. "The pinned Karhunen-Loeve transform of a two dimensional Gauss-Markov field". *Proc. SPIE Conf. Image Process, San Diego, CA.*, (1976).
- [5] A. Z. Meiri and E. Yudilevich. "A pinned sine transform image coder". *IEEE Trans. Commun.*, vol. COM-29, pp. 1728-1753, (1981).
- [6] X. Zhu, A. T. S. Ho and P. Marziliano. "Semi-fragile watermarking authentication and restoration of images using irregular sampling", *Signal Processing: Image Communication*, submitted, (2005).

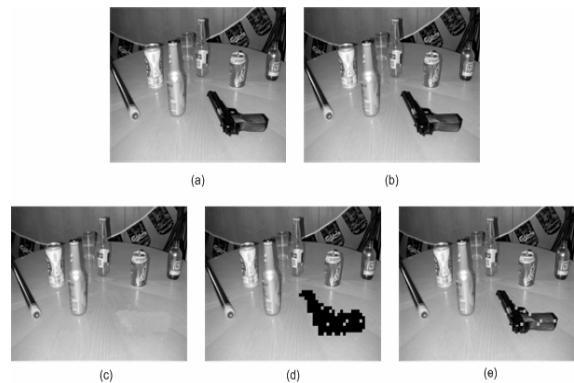


Figure 1: Crime scene analysis: (a) original image; (b) watermarked image; (c) tampered image; (d) authentication result and (e) restoration result

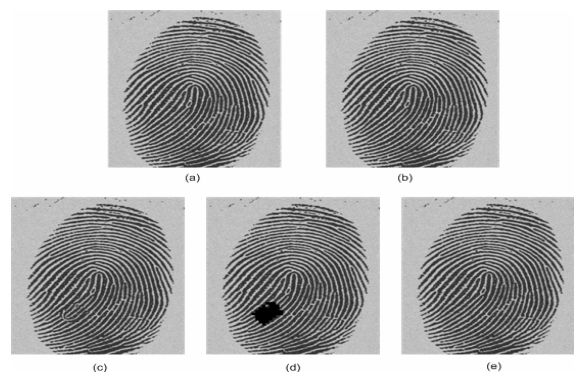


Figure 2: Fingerprint authentication: (a) original image; (b) watermarked image; (c) tampered image; (d) authentication result and (e) restoration result

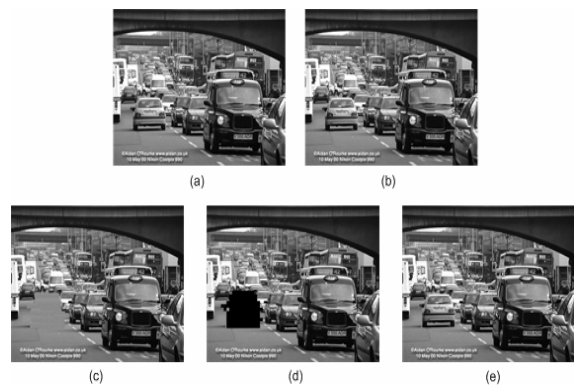


Figure 3: Traffic enforcement: (a) original image; (b) watermarked image; (c) tampered image; (d) authentication result and (e) restoration result