DIGITAL WATERMARKING : AN OVERVIEW

G.Voyatzis, N.Nikolaidis and I.Pitas
Department of Informatics
University of Thessaloniki
Thessaloniki 54006, Greece
Tel,Fax: +3031-996304
e-mail: pitas@zeus.csd.auth.gr

ABSTRACT

In this paper we describe a general framework for image copyright protection through digital watermarking. In particular we present the main features of an efficient watermarking scheme, discuss robustness issues and describe the three main stages of a watermarking algorithm namely watermark generation, embedding and detection.

1 INTRODUCTION

The rapid evolution of digital technology makes the development of reliable and robust schemes for protecting digital still images, audio and video from piracy a matter of urgency. Piracy attacks include illegal access to transmitted data in networks, data content modification and production and retransmission of illegitimate copies [1, 2]. The impact of such attacks might be very large both in financial (profit losses by unauthorised access and use of data) and security terms.

Data transmitted through network communication lines may be protected from unauthorised receivers by applying techniques based on cryptography [3]. Only persons, who possess the appropriate private *key*, can decrypt the received data using a public algorithm implemented either in hardware or in software. Fast implementation of encryption-decryption algorithms is highly desirable.

Data content manipulation can be performed for various legal or illegal purposes (compression, noise removal, malicious data modification). The modified product is not authentic with respect to the original one. Content verification can be performed by attaching *digital signatures* to the transmitted data. A digital signature is an encoded message that matches the content of a particular authentic digital product [3]. Authenticity verification procedures are based on public algorithms and public keys.

Reproduction of a digital product is easy and inexpensive. In a networked environment (like the Word Wide Web) retransmission of copies all over the world is feasible. Copyright ownership can be violated by persons who illegally claim the product exploitation rights. The problem of protecting the intellectual property of digital products has been treated in the last few years with the introduction of the notion of watermarks.

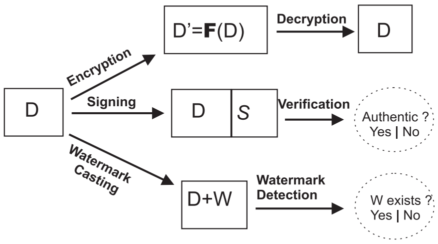

Figure 1: Schematic representation of data encryption, authenticity verification and watermarking.

In this paper we discuss watermark effectiveness in the protection of the intellectual rights on digital products. We will refer, mostly, to watermarking of still digital images. However the concepts introduced in this work can be readily extended to digital audio and video.

2 WATERMARK LITERATURE OVERVIEW

A variety of watermarking techniques has been proposed by various authors in the last three years. The proposed algorithms can be classified in two main classes on the basis of the utilisation of the original image during the detection phase. Algorithms proposed in [11, 12, 9, 10, 20, 13] do not require the original image whereas in those presented in [17, 18, 16, 7] the original image is input in the detection algorithm along with the watermarked image. Detectors of the second type have the advantage to detect the watermarks in images that have been extensively modified in various ways. However detectors of this kind cannot be combined with web-crawling and automatic watermark searching in a digital library.

Watermark embedding can be done either in the spatial domain or in an appropriate transform domain (DCT domain [7, 15, 16, 13], Wavelet transform domain [17, 18], Fourier Mellin domain [20], Fourier Transform domain [19]). In certain algorithms also, the imposed changes take into account the local image characteristics and the properties of the human visual system (perceptual masking) in order to obtain watermarks that are guaranteed to be invisible [15, 12, 13, 17].

3 MAIN FEATURES OF A WATERMARKING SCHEME FOR STILL IMAGES

Watermarks are digital signals that are superimposed on a digital image causing alterations to the original data. A particular watermark belongs exclusively to one owner who is the only person that can proceed to a trustworthy detection of its personal watermark and, thus prove the ownership of the host image. The owner is also the only person that can remove the watermark from the digital data.

Watermarks should possess the following features:

Perceptual Invisibility: The modifications caused by watermark embedding, should not degrade the perceived image quality. However, even hardly visible differences may become apparent when the original image is directly compared to the watermarked one. We therefore make the assumption that the original product is accessible only to the legal owner and such differences remain unnoticed by the observer.

Trustworthy detection: Watermarks should constitute a sufficient and trustworthy proof of ownership on a particular product. Detection false alarms (false positives) should appear extremely rarely (hopefully never). Watermark signals should be characterised by great *complexity*. This is necessary in order to be able to produce an extensive set of sufficiently well distinguishable watermarks. An enormous set of watermarks prevents the recovery of a particular watermark by trial and error procedures.

Associated key: Watermarks should be associated with an identification number so called *watermark key*. The key is used to cast, detect and remove a watermark. Subsequently, the key should be *private* and characterise exclusively the legal owner. Any digital signal, extracted from a digital image, is assumed to be a *valid watermark* if and only if it is associated with a key via a well established algorithm. This condition prevents the creation of *counterfeit* watermarks discussed extensively by Craver et al [4].

Automated detection/search: Watermarks should combine easily with a search procedure that scans any publicly accessible domain in a network environment for illegal deposition of an owner's product.

Statistical invisibility: Watermarks should not be recovered using statistical methods. For example the possession of a great number of digital products, watermarked with the same key, should not dispose the watermark by applying statistical methods. Therefore, watermarks should be image dependent.

Multiple Watermarking: We should be able to embed a sufficient number of different watermarks in the same image. Each watermark should be detectable by using the corresponding unique key. This feature seems necessary because we cannot prevent someone from watermarking an already watermarked image. It is also convenient in cases where the copyright property is transferred from one owner to another (a fingerprinting like process [2]). We mention that the legal image owner is the only one that can dispose a copy containing *only* his/her watermark [5].

Robustness: A digital image can undergo a great deal of different modifications that deliberately (piracy attacks) or not (compression, filtering for noise removal, resizing) affect the embedded watermark. Obviously, a watermark that is to be used as a means of copyright protection should be detectable up to the point that the host image quality remains within acceptable limits. Because of its importance, the watermark robustness issue will be more thoroughly discussed in section 5.

4 WATERMARKING IMPLEMENTATION FUNCTIONS

Let I_o be the original image of size $N \times M$. We can define as *watermark* a 2D digital signal W of the same size having elements:

$$W(i,j) \in \{-1,0,1\} \ , \ 0 \le i < N \ , \ 0 \le j < M \quad (1)$$

A binary or, more generally, a bi-valued form can be also considered. In our definition, zero values denote image pixels or regions that are not affected by the watermarking. In a watermarking scheme one can distinguish three fundamental stages: watermark generation, embedding and detection.

4.1 Watermark generation

Let $\mathcal{W}$ be the set of possible watermark signals. According to the requirement for the existence of an associated key we consider the finite key space $\mathcal{K}$. If $\mathcal{I}$ denotes the set of still digital images, a watermark generation procedure should be defined by the following function :

$$\mathcal{F} : \mathcal{I} \times \mathcal{K} \to \mathcal{W} \ , \ W = \mathcal{F}(I,K) \quad (2)$$

where $K \in \mathcal{K}$ is the watermark key and $I \in \mathcal{I}$ is the image where the watermark will be embedded. For any particular image I and a given watermark signal W the key extraction should be impossible. $\mathcal{F}$ should preferably be a composite function:

$$\mathcal{F} = \mathcal{T} \circ \mathcal{G} \ , \ \mathcal{G} : \mathcal{K} \to \mathcal{W} \ , \ \mathcal{T} : \mathcal{W} \times \mathcal{I} \to \mathcal{W} \quad (3)$$

$\mathcal{G}$ may be a non-invertible pseudo-random number generator having seed the input key K. $\mathcal{T}$ modifies the watermark W produced by $\mathcal{G}$ to obtain a new watermark W' according to the image where the watermarking is applied. We remark that the non-invertibility of $\mathcal{F}$ is

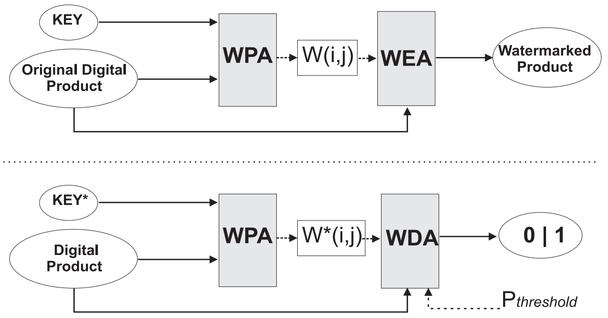

Figure 2: Watermarking Scheme for casting (top) and detection (bottom). WPA, WEA and WDA denote the algorithms for watermark generation embedding and statistical detection

inherited from either $\mathcal{G}$ or $\mathcal{T}$. The watermark modification function $\mathcal{T}$ should take into account only robust image characteristics so that both the original image I_0, the watermarked one I_w and a modified copy of I_w (denoted by I'_w) result in the same watermark :

$$\mathcal{T}(W, I_o) = \mathcal{T}(W, I_w) = \mathcal{T}(W, I'_w) \qquad (4)$$

4.2 Watermark embedding

By considering a watermark $W(i, j)$ produced by $\mathcal{F}$, the embedding process is defined as a superposition of a 2-D digital signal $W(i, j)$ onto the original image $I_o(i, j)$. We denote the embedding procedure by $\mathcal{E}$ and we define it as follows:

$$\mathcal{E} : \mathcal{I} \times \mathcal{W} \times R \rightarrow \mathcal{I} \quad , \quad I_w(i, j) \; = \; I_o(i, j) \oplus L(i, j) W(i, j) \qquad (5)$$

where L is a two dimensional *watermark embedding mask* and $\oplus$ denotes a superposition operator including appropriate truncations and quantisation. The embedding mask L should be image specific and take under consideration the perceiving characteristics of the human visual system.

4.3 Watermark detection

Watermark detection is the most important part in a watermarking algorithm. We denote this procedure by the function $\mathcal{D}$. The detector output may be either a binary (yes/no) decision [11, 12] on the existence of a watermark or a longer bitstream carrying various information [8, 9, 10].

When the watermark is image dependent, the associated key $K \in \mathcal{K}$ is first input in $\mathcal{F}$, W is created and inserted in $\mathcal{D}$. Note that $\mathcal{F}$ should be robust to changes in the image because otherwise it would produce a wrong key when applied on an image that has been manipulated. By taking under consideration the above notions, we define the function $\mathcal{D} \; : \; \mathcal{I} \times \mathcal{K} \rightarrow \{0, 1\}$ as follows:

$$\mathcal{D}(I_w, W) = \mathcal{D}(I_w, \mathcal{F}(I_w, K)) = \begin{cases} 1 & \text{if } W \text{ exists} \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

Detection functions of this kind are the most convenient for creating an efficient watermarking framework for copyright protection. Hypothesis testing [12, 10] or watermark similarity correlators [7, 15] can be used as a basis for such detectors.

The detector output will form a substantial proving evidence of copyright ownership at a law court provided that it is sufficiently trustworthy. This requirement suggests that watermark detection should be a publicly known and globally acceptable procedure. By adopting a hypothesis testing framework, possible errors can be classified in two categories :

Type I error: Watermark is detected although it does not exist in the image. This error is expressed quantitatively by the probability of false alarm (P_{fa}).

Type II error : Watermark is not detected in the image although it exists. Thus, we get an error probability of watermark rejection (P_{rej}).

The total error probability is $P_{err} = P_{fa} + P_{rej}$ and the detection performance increases when P_{err} decreases. However, the reliability of the detection is associated exclusively with the false alarm probability and the following (almost linear) convergence should be satisfied for a randomly selected watermark:

$$\lim_{C(\mathcal{K}) \rightarrow \infty} \; [Prob\{\mathcal{D}(I_W, W) = 1\}] \; = \; 0 \qquad (7)$$

where $C(\mathcal{K})$ denotes the cardinality of the key set. We should mention here that the two types of error compete each other; by decreasing P_{fa}, P_{rej} increases and vice versa.

5 ROBUSTNESS ISSUES

A watermark that is of some practical use should be robust to image modifications up to a certain amount. The most common image manipulations are the following:

Compression. Compression algorithms tend to remove visually insignificant information which is usually where watermarks reside. Some authors propose placing the watermark in the perceptually important components of the image or using watermarks with lowpass characteristics.

Filtering. Attacks based on lowpass filtering (mean, median) can be treated using watermark signals having lowpass characteristics.

Color Quantisation/ Color-Brightness modifications. (histogram modification/equalisation).

Geometric distortions. (scaling, rotation, cropping, deletion or insertion of lines/columns, reflection) To cope with scaling and rotation one can embed the watermark in the Fourier Mellin coefficient space and applying spread spectrum techniques [20]. perform a search within the space of all possible geometric distortions to find the one that has been applied to the image under inspection, insert hidden and secret reference marks, or use correlation based techniques.

Format change. Watermarks are robust to digital image format changes by definition. They are embedded in the main data without producing any effects to format labels.

Although the watermarking algorithms proposed so far cope with some of the above image manipulations, a watermark scheme that copes successfully with all possible attacks has not been proposed yet. Since watermarks should be robust, they are not aimed to used for image authentication [2]. However, watermarks, weakly resistant to modifications, may be used for image content verification [21].

6 CONCLUSIONS

In this paper we have discussed a general watermarking scheme based on public algorithms and a private key encoding-decoding procedure. The watermarks generated by the algorithm should be image dependent to prevent statistical visibility. Furthermore they should be complex in order to provide reliable statistical detection. Third trustworthy parties are not entered in the scheme. The legal owner is the only person who can always provide a copy of the product with only one embedded watermark.

References

[1] B.Macq and J.J.Quisquater, "Cryptology for Digital TV Broadcasting", *Proceeding of the IEEE*, vol 83, no 6, pp.944-957, 1995.

[2] J.F.Delaigle, J.M.Boucqueau, J.J.Quisquater and B.Macq, "Digital images protection techniques in a broadcast framework : an overview", *Proceedings of ECMAST'96*, vol.2, pp 711-727, 1996.

[3] D.R.Stinson, "Cryptography, theory and practice", CRC Press, 1995.

[4] S.Craver, N.Memon, B.L. Yeo and M.M.Yeung, "Can invisible watermarks resolve rightful ownerships?", *IBM Tech.Report* RC20509, 1996.

[5] J.Zhao, "A WWW service to embed and prove digital copyright watermarks", *Proceedings of ECMAST'96*, vol.2, pp 695-709, 1996.

[6] I.J.Cox and M.L.Miller "A review of watermarking and the importance of perceptual modelling", *Proc. of Electronic Imaging'97*, February 1997.

[7] I.J.Cox, J.Kilian, T.Leighton and T.Shammoon "Secure Spread spectrum Watermarking for Multimedia", *Tech. Report*, NEC research Institute, no 95-10, 1995.

[8] C.T.Hsu and J.L.Wu, "Hidden signatures in Images" *Proceedings of ICIP'96*, vol III, pp 223-226, 1996.

[9] G.Voyatzis and I.Pitas, "Applications of Toral automorphisms in image watermarking", *Proceedings of ICIP'96*, vol II, pp. 237-240, 1996.

[10] G. Voyatzis and I. Pitas, "Embedding Robust Logo Watermarks in Digital Images", *Proceedings of DSP'97*, vol 1, pp. 213-216, 1997.

[11] I.Pitas, "A method for signature casting on digital images", *Proceedings of ICIP'96*, vol III, pp 215-218, 1996.

[12] N.Nikolaidis and I.Pitas "Robust image watermarking in the spatial domain", to appear in *Signal Proc.* special issue on *Copyright Protection and Access control*, 1998.

[13] A. Piva, M. Barni, F. Bartolini, V. Cappellini, "DCT-based watermark recovering without resorting to the uncorrupted original image", Proceedings of ICIP'97, Santa Barbara, CA, USA, October 26-29, 1997, Vol I, pp. 520-523.

[14] R. B. Wolfgang and E. J. Delp, "A Watermark for Digital Images," Proceedings of ICIP'96, September 16-19, 1996, Lausanne, Switzerland, pp. 219-222.

[15] Mitchell D. Swanson , Bin Zhu , and Ahmed H. Tewfik , "Transparent Robust Image Watermarking," Proc. of the 1996 IEEE Int. Conf. on Image Processing , Vol. III, PP. 211-214, 1996.

[16] J. O'Ruanaidh, W. Dowling, F. Boland, "Watermarking digital images for copyright protection", IEE Proceedings on Vision, Image and Signal Processing, 143(4), pp 250-256, August 1996.

[17] D. Kundur, D. Hatzinakos, "A Robust Digital Image Watermarking Method using Wavelet-Based Fusion", Proceedings of ICIP'97, Santa Barbara, CA, USA, October 26-29, 1997, Vol I, pp. 544-547.

[18] X.-G. Xia , C. G. Boncelet, and G. R. Arce, " A Multiresolution Watermark for Digital Images" Proceedings of ICIP'97, Santa Barbara, CA, USA, October 26-29, 1997, Vol I, pp. 548-551.

[19] J. O'Ruanaidh, W. Dowling, F. Boland, "Phase watermarking of digital images", *Proc. 1996 IEEE Int. Conference on Image Processing (ICIP 96)*, vol III, pp 239-242.

[20] J. O'Ruanaidh, T. Pun, "Rotation, Scale and Translation Invariant Digital Image Watermarking", Proceedings of ICIP'97, Santa Barbara, CA, USA, October 26-29, 1997, Vol I, pp. 536-539.

[21] M..M. Yeung, F.Mintzer, "An invisible Watermarking technique for image Verification", Proceedings of ICIP'97, Santa Barbara, CA, USA, October 26-29, 1997, Vol II, pp. 680-683.