SEMIFRAGILE HIERARCHICAL WATERMARKING IN A SET THEORETIC FRAMEWORK

Oktay Altun, Gaurav Sharma, Mehmet Celik, Mark Bocko

ECE Dept., University of Rochester, Rochester, NY, USA

ABSTRACT

We introduce a set theoretic framework for watermarking and illustrate its effectiveness by designing a hierarchical semi-fragile watermark that is tolerant to compression and allows tamper localization. Using a quad-tree representation, a spatial resolution hierarchy is established on the image and a watermark is embedded corresponding to each node of the hierarchy. The watermarked image is determined so as to jointly satisfy the multiple constraints of watermark detectability, imperceptibility, and robustness to compression using the method of projections onto convex sets. The spatial hierarchy of watermarks provides a graceful trade-off between robustness and localization under JPEG compression: mild JPEG compression preserves watermarks at all levels of the hierarchy allowing fine localization of malicious changes while aggressive JPEG compression preserves watermarks at coarser levels of the hierarchy still assuring overall image integrity but giving up the capability for localization. Experimental results are presented to illustrate the effectiveness of the method.

1. INTRODUCTION

With the increasing reliance on digital information transfer, methods for the verification of integrity of such information are also continually acquiring increasing importance. Established cryptographic techniques and protocols enable integrity/ownership verification for digital data provided suitable infrastructure is available and the data undergoes no modifications. These methods, however, encounter limitations in the context of multimedia data where one desires resilience to benign signal processing operations such as mild compression and infrastructure constraints make it necessary to embed the integrity/ownership meta-data in the media itself instead of requiring an auxiliary channel. Watermarking (or data-embedding) methods attempt to address these requirements.

In typical watermarking applications, the watermark must satisfy multiple, often conflicting, constraints, common requirements being imperceptibility, detectability, localization capability, and robustness to benign signal processing. Several ad hoc methods and optimization algorithms have been developed for efficient watermark insertion into multimedia under multiple requirements. The optimal transform domain watermark embedding method is a good example of the latter class where the watermark insertion is formulated as a linear programming problem in which the strength of the watermark in frequency domain is maximized subject to a set of constraints in the spatial domain [1]. Though powerful, the formulation is limited to linear constraints, whereas a number of the desirable constraints in watermarking are nonlinear.

This work was supported by the Air Force Research Laboratory/IFEC under grant number F30602-02-1-0129.

In this paper, we propose a set-theoretic framework for watermark embedding. Given multiple requirements , a set theoretic framework is a natural choice to find a solution that satisfies the multiple constraints simultaneously. Set theoretic methods find feasible solutions rather than the optimal solution. Feasibility problems are analytically easier and computationally inexpensive compared to optimization problems and the solutions are acceptable for many engineering applications [2].

We demonstrate the framework by constructing a semi-fragile watermark which is resilient to JPEG compression and also allows tamper localization. The ability to identify regions of suspected alterations and to distinguish those regions from other areas where there is high confidence that the watermarked image has not been damaged is crucial but challenging [3]. Wide use of lossy compression techniques requires detectability of tamper localization of compressed images.

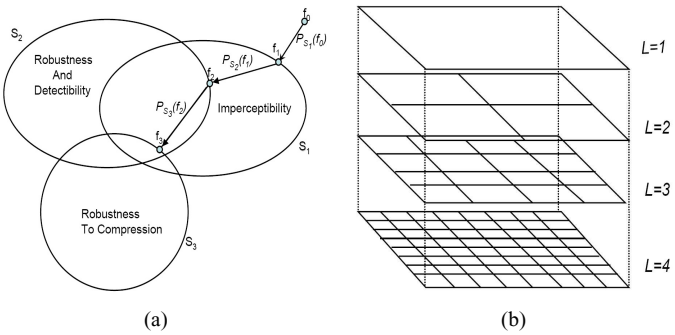
2. SET THEORETIC WATERMARKING AND POCS

The central idea of set theoretic watermarking is to represent each property desired of the watermarked image as a *constraint set*. Thus if there are n desirable properties these are represented as n sets $\{S_i\}_{i=1}^n$, where S_i denotes the set of images that possess the i^{th} property. Any image that lies in the intersection $\bigcap_{i=1}^n S_i$ of all the n constraint sets possesses all the desired properties and may be used as a watermarked version of the image. A practical method for watermarking in the set-theoretic framework then requires techniques for determining an image in the intersection of all constraint sets. The method of projections onto convex sets (POCS) provides a robust algorithm for cases when the sets $\{S_i\}_{i=1}^n$ are all convex [4]. If the intersection set is non-empty, the sequence $\{f_k\}_{k=0}^{\infty}$ generated by successive (relaxed) projections onto the sets converges to a point in the intersection, where

$$f_{k+1} = (T_{S_n}(T_{S_{n-1}}...T_{S_1}(f_k)...)), k = 0, 1, .. \qquad (1)$$

$T_{S_i} = (1 - \lambda_{S_i})I + \lambda_i P_{S_i}, 0 < \lambda_i < 2$ is the relaxed projection operator onto set S_i . For unity relaxation $T_{S_i}(f_k)$ will be equal to $P_{S_i}(f_k)$, which is set to be in the rest of this paper. The initial point f_0 can be arbitrarily be chosen and partly determines the point selected from the intersection.

Figure 1(a) illustrates the general notion of watermark insertion by POCS. The domain is a Euclidean space with dimensions of the cover file where watermark is inserted. Successive projections onto detectability, imperceptibility and robustness to compression provides image adaptive semi-fragile watermark embedding.



(a) (b)

Fig. 1. (a) A generic illustration of semi-fragile watermarking by POCS. (b) Partitioning of an image and the resulting four level hierarchical overlapping block structure.

3. CONSTRAINTS FOR HIERARCHICAL WATERMARKING:

In order to obtain a watermarked signal using POCS we define the following constraints:

3.1. Detectability and robustness

The constraint shown in equation 2 ensures the detectability of the spread spectrum (SS) watermark by the receiver. The resulting image correlated with pseudo-random sequence (generated by the cryptographic key which is available both sides) should be higher than the specified threshold value [5]

$$S_1 \equiv \{X \in \mathbb{R}^{NXM} : \frac{1}{N_h \times M_h}(W_h^* - \overline{W}_h^*)^T \cdot (X_h^* - \overline{X}_h^*) \geq \gamma_h, \forall h\} \quad (2)$$

where over-line represents the mean of the corresponding variable, $N \times M$ is the image dimension, *h* stands for hierarchy and * represents arbitrary permutation of matrix columns to form a vector for conventional illustration of dot product.

The pn-sequence is spectrally shaped by replicating the white noise horizontally, vertically and diagonally. This provides extra robustness to compression since watermark is designed not to be at high pass region which is very likely to be removed by compression or malicious attacks. Variable R stands for how many times the pn-sequence is replicated. Figure 1(b) illustrates partitioning of an image and the resulting multilevel hierarchical overlapping block structure.

3.2. Point-wise Fidelity

The visual fidelity of the image is guaranteed by two visual constraints. The first one is imposed by a spatial domain texture masking model. Voloshynoskiy et al. has proposed this visual model which outputs allowable distortion at pixel level given the original image [6]. In this particular texture masking model, image is modelled as the sum of the local mean and an error term, the latter of which is modeled by a generalized Gaussian distribution. Equation 3 shows this constraint where X is an image size matrix which lies in the allowable lower (L) and upper (U) bounds.

The resulting constraint can be expressed as follows:

$$S_2 \equiv \{X \in \mathbb{R}^{NXM} : L \leq X \leq U\} \quad (3)$$

3.3. Overall Fidelity

The second visual model is proposed by Mannos et al [7]. They have proposed a visual distortion metric for monochrome images which takes into account of the fact that human observer is more sensitive to some spatial frequencies than others and he is more sensitive to intensity errors in gray regions than white. He formulates a parametric filter based on experimental results which we use in our method. The second constraint is formulated as an inequality that forces the overall visual distortion metric of the image be smaller then some threshold value. $\| \cdot \|$ represents Euclidean distance , H_w is the filter in transform domain and $X_{w,0}$ represents the original image in transform domain.

$$S_3 \equiv \{X \in \mathbb{C}^{NXM} :\| H_w \cdot X_w - H_w \cdot X_{w,0} \| \leq \theta\} \quad (4)$$

3.4. Robustness To Compression

The requirement of robustness to compression can be roughly expressed by the following mathematical expression:

$$S_4 \equiv \{X \in \mathbb{R}^{NXM} : \frac{1}{N \times M} W^* \cdot IDCT(Q[DCT(X^*)]) \geq \gamma\} \quad (5)$$

where Q stands for JPEG quantization scheme , DCT and IDCT represents discrete cosine transform and inverse discrete cosine transform respectively. However, this constraint is not convex.

So we approximated it with the following equation:

$$\widehat{S}_4 \equiv \{X \in \mathbb{R}^{NXM} : \frac{1}{N \times M} W^* \cdot IDCT(Q_{NZ(X_0)}[DCT(X^*)]) \geq \gamma\} \quad (6)$$

where $Q_{NZ(X_0)}$ refers to the non-zero quantized DCT coefficients of the original image. This approximation has the underlying assumption that the DCT coefficients that is quantized to zero after compression is causing the major loss of watermark information.

4. EXPERIMENTAL RESULTS

The detector response of pn-sequence with hierarchically watermarked Goldhill image is illustrated in figure 2. The method shows significant performance in cancelling all interference at all levels by bringing the detector response on unity. Among images Goldhill, Lena, Barbara, Mandrill, Washat, Peppers, Boat and Zelda; all hierarchically watermarked images show similar histogram except Zelda , Boat and Lena at layer 4. Since these images have relatively flat regions in certain areas, there are slight deviations from unity detector response at layer 4.

The visual fidelity of the resulting image is controlled by three parameters : γ_h , θ, P_0 , P_1 and Q_R . These parameters relax/tighten the sets of visual fidelity. Tightening the constraints

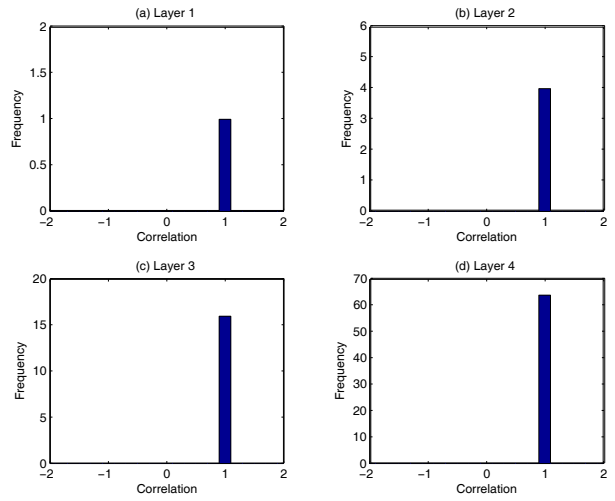


Fig. 2. Histogram of detector responses of watermark sequence with hierarchically watermarked uncompressed Goldhill image (Mandrill, Washsat, Barb, Peppers give same response as well). Each layer gives a unity detector response.

increases the quality of the resulting image however tightening beyond some value will cause the set move far from the other sets leaving no room for intersection. Relaxing the allowable distortion sets , on the other hand, will decrease the image fidelity.

We restricted the total number of iterations to be 60. However, it is observed that usually 30 iterations is enough for convergence to a very close point on the intersection of the sets.

The watermark's resistance to compression is tabulated in table 1. The watermarks vanish consistent with hierarchy with increasing compression levels. We have chosen the threshold level for detector as 0.6 at all levels of the hierarchy.

The watermarked image was tampered to illustrate the tamper detection capability of the hierarchical watermark as shown in figure 5 . The tampered region is illustrated by shading and numbering the regions. The image is compressed by Q=40 and the watermark is detected when threshold is 0.6. Contrary to the authentication schemes, the scheme distinguishes the regions from other areas where there is high confidence that the watermarked image has not been damaged.

5. CONCLUSION

In this paper, we introduce a set-theoretic framework for watermarking that naturally allows multiple requirements to be imposed on the watermarked image. We demonstrate the usefulness of the framework by developing a hierarchical semi-fragile watermark that incorporates constraints of detectability, robustness against JPEG compression, and imperceptibility. These requirements are formulated as convex sets and a watermarked image is computed using the method of projections onto convex sets (POCS). Combined with the hierarchical nature of the watermark the constraints ensure the capability for tamper localization along with robustness against JPEG compression. Simulation results demonstrate that the method meets its design objectives and that the framework provides a useful method for meeting the often conflicting requirements in watermarking applications.

Fig. 3. Watermarked Washsat image. ($P_0 = 30$, $P_1 = 3$, Q=30, L=4, R=2 and γ=1. PSNR=30.53 dB)

Fig. 4. Manipulated image. Downtown is replicated.

6. APPENDIX A

The projections are performed by using Lagrange multipliers.

$$X_{i+1} = P_j(X_i) = \arg\min_{X \in S_j} \|X - X_i\| \quad (7)$$

The Lagrangian function of the first projection can be written as follows :

$$L_1 = \sum_i \sum_j [X(i,j)X_i(i,j)]^2 - \lambda \cdot \sum_i \sum_j [W(i,j) - \overline{W}][X(i,j) - \overline{X}] \quad (8)$$

The square root over the objective function is dropped due to monotonicity. Taking the derivative of the Lagrangian w.r.t. X(i,j)

	$Q=90$	$Q=80$	$Q=70$	$Q=60$	$Q=50$	$Q=40$	$Q=30$	$Q=20$	$Q=10$
$1^{st}lyr$	8/8	8/8	8/8	8/8	8/8	8/8	8/8	8/8	2/8
$2^{nd}lyr$	32/32	32/32	32/32	32/32	32/32	32/32	31/32	27/32	11/32
$3^{rd}lyr$	128/128	128/128	128/128	128/128	128/128	125/128	123/128	106/128	46/128
$4^{th}lyr$	512/512	512/512	509/512	507/512	503/512	482/512	453/512	370/512	207/512

Table 1. Detection of authentication watermark at various JPEG compression levels . Detection threshold is chosen to be 0.60 The experiment is performed on 8 different images and R=2.

Fig. 5. Illustration of tamper localization by hierarchical semi fragile watermark when image is JPEG compressed with Q=40.

and λ gives first order conditions (FOC1) and (FOC2) respectively. This formulation is valid for each hierarchy level.

$$FOC1 : X(i,j) = X_i(i,j) - \lambda \frac{(N+1)}{2N}[W(i,j) - \overline{W}] \quad (9)$$

$$FOC2 : \sum_i \sum_j [W(i,j) - \overline{W}][X(i,j) - \overline{X}] = \gamma \quad (10)$$

substituting X(i,j) into FOC2 and solving for λ gives the value of Lagrange multiplier. Substituting non-negative λ into FOC1 reveals the projection.

Projection onto overall fidelity constraint can be achieved by forming the Lagrangian and taking the derivative of the Lagrangian w.r.t. X(i,j) and λ.

$$FOC3 : X_w(i,j) - X_w^i(i,j) = \lambda \cdot [X_w(i,j) \cdot H_w(i,j) - X_{w,o}(i,j) \cdot H_w(i,j)] \quad \forall i,j \quad (11)$$

$$FOC4 : \sum_j \sum_i [(X_w(i,j) \cdot H_w(i,j) - X_{w,o}(i,j) \cdot H_w(i,j)]^2 = \theta \quad \forall i,j \quad (12)$$

Solving FOC3 and FOC4 simultaneously will complete the projection onto overall fidelity constraint.

Projection onto point-wise fidelity has an immediate explicit solution:

if $\lambda > 0$ and $X(i,j) > U(i,j)$ then

$$X(i,j) = U(i,j) \quad \forall i,j \quad (13)$$

if $\lambda > 0$ and $X(i,j) < L(i.j)$ then

$$X(i,j) = L(i,j) \quad \forall i,j \quad (14)$$

The projection onto robustness to compression is performed by the following algorithm:

Algorithm:
- Find the DCT coefficients of original image.
- Find the DCT coefficients that gives 0 *after compression.*
- Replace the DCT coefficients that is quantized to zero after compression with the original DCT coefficients.

7. REFERENCES

[1] S. Pereira, S. Voloshynoskiy, and T. Pun, "Optimal transform domain watermark embedding via linear programming," *Signal Processing*, vol. 81, no. 6, pp. 1251–1260, Jun. 2001.

[2] K. C. Haddad, H. Stark, and N. P. Galatsanos, "Constrained fir filter design by the method of vector space projections," *IEEE Transactions On Circuits And SystemsII: Analog And Digital Signal Processing,*, vol. 47, no. 8, pp. 714–725, Aug. 2000.

[3] E. J. Delp, C. I. Podilchuk, and E. T. Lin, "Detection of image alterations using semi-fragile watermarks," in *SPIE International Conf. on Security and Watermarking of Multimedia Contents II, No. 14, EI '00, San Jose, USA, Jan 2000.*

[4] P. L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, Feb. 1993.

[5] I. J. Cox, M. L. Miller, and J. A. Bloom, "Digital watermarking," *Morgan Kaufmann*, Jul. 2001.

[6] S. Voloshynoskiy, A. Herrigel, N. Baumgaertner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," *Proceedings of the Third International Workshop on Information Hiding,Dresden, Germany,1999*.

[7] J. L. Mannos and D. L. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, Jul. 1974.