A Set Theoretic Framework for Watermarking and Its Application to Semifragile Tamper Detection

Oktay Altun, *Student Member, IEEE*, Gaurav Sharma, *Senior Member, IEEE*, Mehmet U. Celik, *Member, IEEE*, and Mark F. Bocko, *Member, IEEE*

***Abstract*—We introduce a set theoretic framework for watermarking. Multiple requirements, such as watermark embedding strength, imperceptibility, robustness to benign signal processing, and fragility under malicious attacks are described as constraint sets and a watermarked image is determined as a feasible solution satisfying these constraints. We illustrate that several constraints can be formulated as convex sets and develop a watermarking algorithm based on the method of projections onto convex sets. The framework allows flexible incorporation of different constraints, including embedding strength requirements for multiple watermarks that share the same spatial context and different imperceptibility requirements based on frequency-weighted error and local texture perceptual models. We illustrate the effectiveness of the framework by designing a hierarchical semifragile watermark that is tolerant to mild compression, allows tamper localization, and is fragile under aggressive compression. 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 spatial hierarchy of watermarks provides a graceful tradeoff between robustness and localization under mild JPEG compression, where watermarks at coarser levels demonstrate progressively higher immunity to JPEG compression. Under aggressive compression, watermarks at all hierarchy levels vanish, indicating a lack of trust in the image data. The constraints implicitly partition watermark power in the resolution hierarchy as well as among image regions based on robustness and invisibility requirements. Experimental results illustrate the flexibility and effectiveness of the method.**

***Index Terms*—Projections onto convex sets (POCS), robustness to compression, semifragile watermark, set theoretic watermarking, spread-spectrum watermark, tamper localization.**

I. Introduction

THE widespread distribution capability afforded by the Internet and the ready availability of image manipulation tools have brought to the forefront several security concerns with the use of digital images. Digital processing and the communications infrastructure, however, also allow us to address these concerns using a variety of techniques ranging from conventional cryptographic methods [1] to digital forensics [2] and watermarking [3]. The choice among these different options is often dependent on the specific application and the operating constraints. In a number of multimedia security applications, watermarks are particularly attractive because they are embedded within the content. This allows them to be readily incorporated in existing infrastructure and also permits content-related functionality, such as localization and tolerance of benign signal processing to be easily built-in.

In typical watermarking applications, the watermark must satisfy multiple, often conflicting constraints, with the common requirements being imperceptibility, detectability, localization capability, robustness to signal processing, and fragility under malicious attacks. The exact bounds of these requirements depend on the specific needs of the application. Copyright protection and content tracking need a noise-tolerant watermark design [5], [6] while authentication applications require fragility of the watermark even when minute changes are performed on the multimedia [7]–[11]. On the other hand, semifragile watermarks should tolerate content-preserving (nonmalicious) lossy transformations (e.g., compression), but should detect malicious manipulations (e.g., removal of objects from a scene [12]–[14]).

A number of ad-hoc methods and optimization-based 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. 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 [24]. Though powerful, the formulation is limited to linear constraints, whereas a number of the desirable constraints in watermarking are nonlinear.

In this paper, we propose a set theoretic framework for watermarking. The framework is a natural choice for finding a solution that simultaneously satisfies the multiple constraints in watermarking. Unlike optimization methods that find a solution that maximizes (or minimizes) an objective function, set theoretic methods find feasible solutions that satisfy required constraints but do not necessarily satisfy an optimality criterion. The power of set theoretic methods comes from their capability to incorporate more of the constraints and more realistic constraints, which are sometimes ignored in optimization frameworks. In addition, feasibility problems are often analytically easier and computationally inexpensive compared to the optimization problems and the solutions are acceptable for many engineering problems [16]–[20].

We show that several requirements in watermarking applications can be mapped to convex constraints or can be closely approximated by convex constraints. These include watermark de-

Manuscript received August 8, 2005; revised June 3, 2006 This work was supported by the Air Force Research Laboratory/IFEC under Grant F30602-02-1-0129. Parts of this work were presented at IEEE ICIP'05 and EUSIPCO'05. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ingemar Cox.

H. O. Altun and M. F. Bocko are with the Electrical and Computer Engineering Department, University of Rochester, Rochester, NY 14627-0231 USA (e-mail: altun@ece.rochester.edu; bocko@ece.rochester.edu).

G. Sharma is with the Electrical and Computer Engineering Department and also with the Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14627-0126 USA (e-mail: gaurav.sharma@rochester.edu).

M. U. Celik is with Information and System Security Group, Philips Research Europe, Eindhoven 5656 AA, The Netherlands.

Digital Object Identifier 10.1109/TIFS.2006.885018

tectability, noise robustness, multiple watermark detectability, imperceptibility, robustness to lossy compression, and fragility under aggressive compression. This allows determination of the feasible solution using the powerful method of projections onto convex sets (POCS).

We illustrate the utility and flexibility of the framework by applying it to the problem of semifragile watermarking. The widespread use of lossy compression provides motivation for watermarking techniques that are tolerant to compression while still allowing identification of regions with suspected alterations [21]. However, in a number of sensitive medical and military applications where perceptual integrity of even fine spatial detail has utmost importance, aggressive lossy compression can be considered as a malicious attack that causes loss of vital information. In such a scenario, it is desirable that watermarks: 1) are robust against mild compression; 2) distinguish manipulated regions from other areas; and 3) vanish under aggressive compression to indicate the loss of significant visual content.

We demonstrate that the framework allows considerable freedom in designing the semifragile watermarked image to meet the constraints of imperceptibility, robustness, and detectability in the presence of mild compression, while simultaneously providing localization capability. The semifragile spread-spectrum hierarchical watermark that we present provides a graceful tradeoff 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 removes watermarks at all hierarchy levels.

The rest of this paper is organized as follows. In Section II, we introduce the set theoretical framework for watermarking, suggesting possible constraints and applications. In Section III, we describe a semifragile hierarchical multiple layer watermarking technique implemented in the framework. In Section IV, we present an experimental evaluation of the proposed semifragile watermarking method and discuss the characteristics of the algorithm, particularly its capability to implicitly manage the distribution of power to meet constraints of imperceptibility and robustness to compression. Section V presents concluding remarks. Mathematical details of the visual texture model incorporated within the method and projections are presented in Appendices A and B, respectively.

II. Set Theoretic Framework for Watermarking

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 ith property. Any image that lies in the intersection $\bigcap_{i=1}^n S_i$ of all the n constraint sets possesses all of the desired properties and may be used as a watermarked version of the image.

A. Watermark Insertion by POCS

A practical method for watermarking in the set-theoretic framework requires techniques for finding an image in the intersection of all constraint sets. The method of projections onto convex sets [15]–[17] provides a robust algorithm for cases when the sets $\{S_i\}_{i=1}^n$ are all convex.

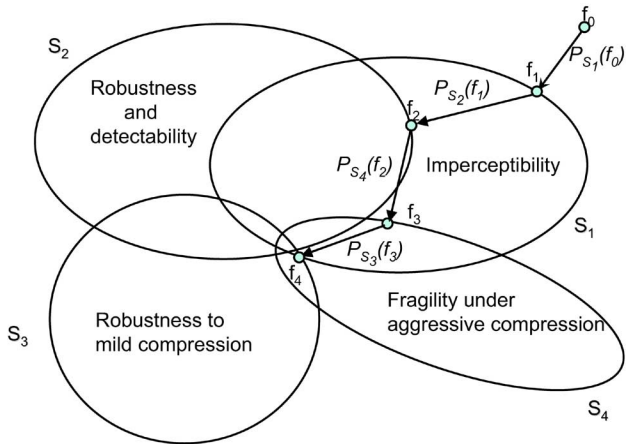

Fig. 1. Schematic illustration of semifragile watermarking by POCS.

POCS Algorithm Statement:

1) *Initialization:* $k \leftarrow 0$, $f_0 \leftarrow$ arbitrary image.
2) *Projection step*: Project sequentially onto convex constraint sets $\{S_j\}_{j=1}^n$ to obtain the next iterate, that is

$$f_{k+1} \leftarrow T_{S_n}(\ldots T_{S_2}(T_{S_1}(f_k))\ldots) \tag{1}$$

where $T_{S_i} = (1 - \lambda_{S_i})I + \lambda_i P_{S_i}, 0 < \lambda_i < 2$ is the relaxed projection operator onto set S_i, with $P_{S_i}(f) = \arg\min_{g \in S_i} \|g - f\|$ denoting the projection of f onto the set S_i. For unity relaxation, $T_{S_i}(f_k) = P_{S_i}(f_k)$, which will be the case considered throughout this paper.

3) *Convergence check*: If $f_{k+1} \in (\bigcap_j S_j)$, set watermarked image as f_{k+1} and terminate iterations, else set $k \leftarrow k+1$ and go to 2).

Fig. 1 schematically illustrates the process of watermark insertion by POCS for a semifragile watermarking scenario. The "points" $f_0, f_1, \ldots, f_n, \ldots$ in this figure represent images with dimensions identical to the dimensions of the cover image in which the watermark is to be inserted. If the intersection set is nonempty, the sequence $\{f_k\}_{k=1}^{\infty}$ generated by the algorithm is guaranteed to converge to a point in the intersection [15], providing an image satisfying all requisite constraints for the semifragile watermark. We define these underlying constraint sets in the next section.

Note that in the set-theoretic embedding framework, the knowledge of the cover data is utilized in formulating the constraint sets. The method therefore is an instance of "informed embedding" [22, p. 132] and is preferable to "blind embedding" methods where the cover data are treated purely as noise. The proposed method, however, is not an "informed coding" technique, typified by quantization index modulation [6].

B. Constraints for Set Theoretic Watermarking

1) Watermark Detectability/Strength of Spread Spectrum Embedding: Spread spectrum embedding [5] is a common watermark embedding technique. Conventionally, the embedding process consists of the addition of a key-dependent pseudo-noise sequence to the cover image. The detection is performed by correlating a test image against the same pseudo-noise sequence and comparing the result against a

threshold to determine whether a watermark is present or not. In the set-theoretic framework, instead of defining an explicit watermark insertion technique, we implicitly insert the watermark by imposing a constraint on the (watermarked) image for exceeding a preset minimum embedding (detection) strength.

For our detector, we adopt the mean corrected linear correlation metric, which provides robustness against additive noise and resilience to valumetric scaling (of images) [22, p. 127]. The constraint on watermark embedding strength is formulated by imposing a minimal value γ for the correlation. If $X^* \in R^{M\times N}$ denotes the image with dimensions $M \times N$ and $W^* \in R^{M\times N}$ denotes the pseudonoise watermark sequence, we denote $X = \mathrm{vec}(X^*) \in R^{MN}$ and $W_0 = \mathrm{vec}(W^*) \in R^{MN}$ as the vectors obtained by stacking together the columns of each. We will adopt the notation in terms of 1-D vectors throughout and assume that any image operators are also represented as matrices/functions conforming with the vector representation. The watermark detectability constraint set is then given by

$$S_1 \equiv \{X : (W_0 - \overline{W_0})^T (X - \bar{X}) \geq \gamma\} \quad (2)$$

$$= \{X : W^T (X - \bar{X}) \geq \gamma\} \quad (3)$$

where $W = W_0 - \overline{W_0}$ and $\bar{Z}$ represent a vector of the same size as Z and having its elements as the sample mean of the vector Z.[1] Unless indicated otherwise, we also adopt the same notational convention for subsequent constraint set definitions. Note that in the formulation above, we are assuming that the detector does not have access to the original image; hence it performs blind detection and for nonblind scenarios, the constraint is readily modified.

The definition of (3) assumes watermark embedding in the spatial domain. The proposed method, however, is generic and can be applied to most transform domains which may be attractive for their properties. For any linear transformation $\mathcal{T}_\mathcal{L}$ of the image, assuming detection in the transform domain, the detectability constraint becomes

$$S_1 \equiv \{X : W^T (\mathcal{T}_\mathcal{L}(X) - \overline{\mathcal{T}_\mathcal{L}(X)}) \geq \gamma\}. \quad (4)$$

Thus, the watermark detectability constraint is readily extended to watermarking in various domains, such as the discrete-cosine transform, fractional-Fourier transform [23], and wavelet transform.

It is worth noting a few of the advantageous properties of the implicit watermark embedding technique proposed here. Since the constraint is formulated in terms of the response of the correlation detector, the interference of the host signal is automatically taken into account and the method does not require adjustments for different images. In this sense, the method utilizes knowledge of the cover image and is superior to additive watermark insertion. The constraint, however, does not guarantee invisibility of the embedded watermark. Another noteworthy aspect is that the method also does not mandate a specific choice of the pseudonoise sequence $W^* \in R^{M\times N}$. In particular, the sequence may not be white and may be spectrally shaped to improve performance in the presence of attacks as suggested by other works [28], [29].

[1]Typically, the embedding strength in the left-hand side of the inequality in (3) would be normalized for the image size, by multiplication with $(1/MN)$; for succinct notation, we will find it convenient to absorb the factor MN into the upper bound for this and subsequent sets.

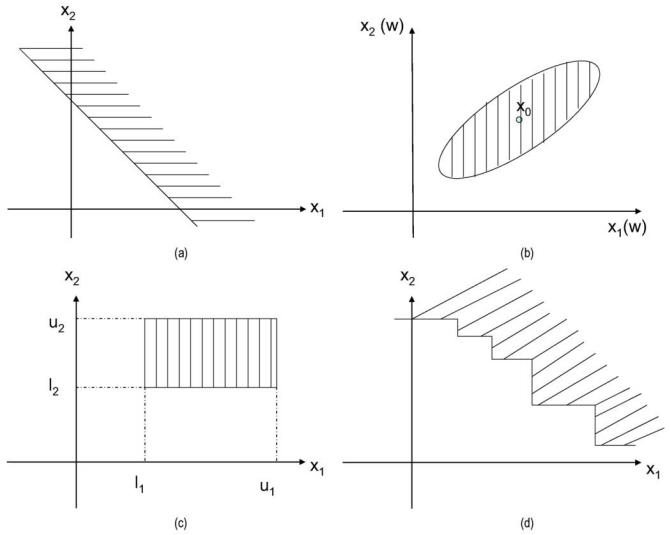

Fig. 2. Schematic illustration of constraint sets in 2-D representation. (a) Watermark detectability. (b) Overall fidelity. (c) Pixel-wise visual fidelity. (d) Robustness to quantization.

Multiple Watermark Embedding: In applications where multiple watermarks are embedded within an image, each individual watermark will encounter interference from others (i.e., the detector response to each watermark is weakened by the embedding of subsequent watermarks). The interference problem is implicitly handled within the set-theoretic watermarking framework, by introducing a separate constraint for each watermark. The image that satisfies all constraints will bear each watermark with sufficient strength. Thus, if we have K watermarks $\{W_j\}_{j=0}^{K-1}$ that are to be embedded in the same image, there are K corresponding constraint sets

$$S_1^j \equiv \left\{X : W_j^T (X - \bar{X}) \geq \gamma_j\right\}, \quad j = 0, \ldots, K-1. \quad (5)$$

Note that the K watermark detectability constraints in (5) are simultaneously imposed on the watermarked image. Therefore, in the multiple watermarking scenario, the set theoretic framework accounts for any interwatermark interference present at the receiver. The formulation allows for the simultaneous insertion of multiple watermarks or sequential insertion, provided that at each stage, the previously embedded watermarks are known, for instance, for a multibit embedding technique [23].

The watermark detectability constraint sets are clearly convex and are schematically shown in Fig. 2(a).

2) Watermark Imperceptibility: The watermarked image should be perceptually (almost) identical to the original image. We incorporate this requirement through the introduction of two image fidelity constraints: an overall visual fidelity constraint formulated in terms of a linear visual system model and a pixel-wise image-fidelity constraint determined in terms of a noise visibility function.

Overall Image Fidelity Constraint: The sensitivity of human observers to changes introduced in an image exhibits a strong spatial frequency dependence. Typically, changes in

image content at low spatial frequencies are perceived more readily than changes at high frequencies. This behavior can be represented, to a first order, by assuming that the perceived image in response to an input image X is given by HX, where H represents a 2-D spatial low-pass filter.[2] Using this approximation, a constraint on overall image fidelity can be formulated by requiring that the difference between the perceived watermarked image HX and the perception of the original image HX_0 should be small (where X_0 denotes the original image). Using the Euclidean norm for quantifying the difference, the constraint therefore becomes

$$S_2 \equiv \{X : \|HX - HX_0\| \leq \theta\} \tag{6}$$

where $\|v\|$ represents the Euclidean norm of v, and θ is a suitably chosen threshold.

The spatial filter H is determined by the specific visual system model employed. In particular, we employ the model proposed by Mannos *et al.* [26], which has been extensively utilized in image processing research. For the model, the spatial filter H is represented in the frequency domain by a radially isotropic function $H(f_r) = 2.6[0.0192+0.114f_r]e^{-(0.114f_r)^{1.1}}$ where f_r denotes the radial frequency in cycles per degree [26]. It is readily seen that the constraint in (6) is convex and a schematic 2-D geometric representation is illustrated in Fig. 2(b).

Pixel-Wise Image Fidelity Constraint: Since the overall fidelity constraint (6) is based primarily on psychophysical data for individual sinusoidal stimuli, it does not adequately handle localized perturbations of the image in a small area. Therefore, we use an additional model to limit local perturbations to ensure imperceptibility. Our model exploits the perceptual phenomenon of spatial masking [27] in which perturbations introduced in an image region at a frequency are masked by stronger image content at similar frequencies. In particular, we use the spatial-domain texture masking model proposed by Pereira *et al.* [24]. Given an original image, the model predicts the allowable distortion at each pixel level that is visually tolerable, leading, in turn, to pixel-wise upper and lower bounds for the difference from the original image. The resulting constraint can be expressed as

$$S_3 \equiv \{X : D_L \leq (X - X_0) \leq D_U\} \tag{7}$$

where D_U and D_L are the same size as X and represent the pixel-wise upper and lower bounds on the distortion, X_0 is the original image, and the inequalities in (7) apply termwise. Additional relevant details of the model are summarized in Appendix A. The constraint can alternately be expressed as $S_2 = \{X : U \leq X \leq L\}$ where $U = X_0 + D_U$ and $L = X_0 + L_D$ form for pixelwise upper and lower bounds. The constraint is obviously convex and is illustrated for the 2-D case in Fig. 2(c).

Note that the imperceptibility constraints apply to the watermarked image and are therefore directly applicable in both single and multiple watermark embedding scenarios. This is in contrast with explicit embedding methods, where the power distribution among the multiple watermarks must be carefully controlled in order to meet this constraint.

3) Robustness to Compression Constraint: Resilience against nonmalicious changes is a desirable property in several applications of semifragile and robust watermarks, which is extremely challenging because of the large class of nonmalicious changes that may be introduced [22]. Here, we concentrate on one important class of alterations consisting of lossy compression. Specifically, we focus on transform coding techniques, for which the lossy component consists of a quantization operation in the transform domain.[3] Resilience of the watermarked image to compression can be achieved by requiring that the detector response to compressed versions of the image exceeds a desired threshold γ. This requirement yields the constraint set

$$S_4 \equiv \{X : W^T(\mathcal{T}_{\mathcal{I}}(Q[\mathcal{T}_{\mathcal{F}}(X)]) - \overline{\mathcal{T}_{\mathcal{I}}(Q[\mathcal{T}_{\mathcal{F}}(X)])}) \geq \gamma\} \tag{8}$$

where $Q[\cdot]$ denotes the quantizer, $\mathcal{T}_{\mathcal{F}}$ represents the transform operation from the spatial domain into the transform domain, and $\mathcal{T}_{\mathcal{I}}$ represents the inverse transform. We assume that the transform $\mathcal{T}_{\mathcal{F}}$ is linear and invertible so that $\mathcal{T}_{\mathcal{F}}(X) \in R^{1\times MN}$ and the quantizer $Q[\cdot]$ operates on a term-by-term basis so that it is the direct product of the scalar quantizers $\{Q^k[\cdot]\}_{k=0}^{MN-1}$, where $Q^k[\cdot]$ is the quantizer for the kth transform coefficient. In the case of JPEG, for example, $\mathcal{T}_{\mathcal{F}}$ is the discrete cosine transform (DCT) and $\mathcal{T}_{\mathcal{I}}$ is the inverse discrete cosine transform (IDCT), and the quantizer $Q[\cdot]$ then consists of the 64 quantizers for the DCT coefficients within each 8×8 block of DCT coefficients (where the frequency-dependent scaling factor is included as part of the quantizer). The constraint of (8) is usually not convex. A typical 2-D representation illustrating this fact is shown in Fig. 2(d). Motivated by the observation that typical transform coding schemes provide coding gain through the compaction of signal energy into a few coefficients, we approximate the set (8) by the following set:

$$\hat{S}_4 \equiv \{X : W^T(\mathcal{T}_{\mathcal{I}}(Q_0[\mathcal{T}_{\mathcal{F}}(X)]) - \overline{\mathcal{T}_{\mathcal{I}}(Q_0[\mathcal{T}_{\mathcal{F}}(X)])}) \geq \gamma\} \tag{9}$$

where $Q_0[\cdot]$ refers to the function determined from the original image X_0 by defining its constituent scalar "quantizer" functions as

$$Q_0^k[t] = \begin{cases} 0, & Q^k[(\mathcal{T}_{\mathcal{F}}(X_0))_k] = 0 \\ t, & \text{otherwise;} \end{cases} \quad k = 0, 1, \ldots, MN-1. \tag{10}$$

where $(\mathcal{T}_{\mathcal{F}}(X_0))_k$ denotes the kth transform coefficient of the original image X_0. Thus, the function $Q_0[\cdot]$ sets transform coefficients that are zero in $Q[\mathcal{T}_{\mathcal{F}}(X_0)]$ to zero and leaves other coefficients unchanged. This approximation has the underlying assumption that the transform coefficients that are quantized to zero cause the major loss of watermark information. We can readily see that for our definition of $Q_0[\cdot]$, we have $Q_0(Z+Y) =$

[2]The visual system also includes significant point-wise nonlinearity. Common digital image representations, however, already include a compensation for this nonlinearity [25] and its effect can therefore be ignored with minimal error.

[3]The JPEG compression standard is the prime and most commonly used example of transform coding that we utilize in our implementation.

$Q_0(Z) + Q_0(Y)$. Hence, from the linearity of the transformation $\mathcal{T}_\mathcal{F}$ the convexity of the set in (9) follows immediately.

4) Fragility Under Aggressive Compression Constraint: While compression is a content preserving benign signal processing operation, under aggressive compression, there can be significant loss of perceptual data. Therefore, in semifragile watermarking, it is sometimes desirable that watermarks disappear under aggressive compression. This fragility of the watermarked image to aggressive compression can be achieved by requiring that the detector response to highly compressed versions of the image be below the detection threshold γ. This requirement provides the constraint set

$$S_5 \equiv \{X : W^T(\mathcal{T}_\mathcal{I}(Q^A[\mathcal{T}_\mathcal{F}(X)]) - \overline{\mathcal{T}_\mathcal{I}(Q^A[\mathcal{T}_\mathcal{F}(X)])}) \leq \gamma\} \tag{11}$$

where $Q^A[\cdot]$ denotes the quantizer for aggressive compression and other terms are as defined in (8). Just like the robustness to compression constraint, this set is not convex; however, arguing as we did before, a convex approximation is obtained as

$$\hat{S}_5 \equiv \left\{X : W^T\left(\mathcal{T}_\mathcal{I}\left(Q_0^A[\mathcal{T}_\mathcal{F}(X)]\right) - \overline{\mathcal{T}_\mathcal{I}\left(Q_0^A[\mathcal{T}_\mathcal{F}(X)]\right)}\right) \leq \gamma\right\} \tag{12}$$

where the function $Q_0^A[\cdot]$ is defined as in (10) with Q^k being replaced by the corresponding aggressive compression quantizer Q_A^k (for the kth transform coefficient).

III. Semifragile Hierarchical Watermarking in a Set Theoretic Framework

Using the constraint sets defined in the preceding section, the method of POCS may be utilized to embed a watermark as outlined in Section II-A. We illustrate the flexibility of the framework with a semifragile watermarking scheme that utilizes multiple watermarks for localization while maintaining a desired level of imperceptibility and tolerance to compression.

Semifragile watermarks can be designed by carefully tuning a robust watermark so that it is removed if the distortion exceeds a particular level [22]. We chose spread-spectrum embedding for this purpose. Noise immunity and reliable detection in spread-spectrum watermarking require long (spreading) sequences. Short sequences, while providing good localization accuracy, are sensitive to noise and may fail to provide sufficient immunity to compression [21]. There is an inherent conflict between the capability to localize malicious changes within the image (short sequences) and resilience against lossy compression (long sequences)—both desirable characteristics of semifragile watermarks. We, therefore, use spread-spectrum embedding in a hierarchical manner which ensures better localization and less immunity to lossy compression at ~~low~~ lower levels and less localization but better immunity at ~~high~~ higher levels.

Our hierarchical block-based watermarking technique adopts a multilevel spatial hierarchy previously used in [11]. A partition of the image into nonoverlapping blocks constitutes the highest level of the hierarchy. Successive levels of the hierarchy are formed by combining distinct groups of blocks at a preceding level of the hierarchy. In general, the number of blocks from a level of the hierarchy that are combined to form a block at the next lower level of the hierarchy may be arbitrarily chosen; however, in order to keep the notation and the description simpler, we assume for the rest of this paper that the region of 2×2 blocks at a given level of the hierarchy be combined to create a block at the next lower level of the hierarchy.

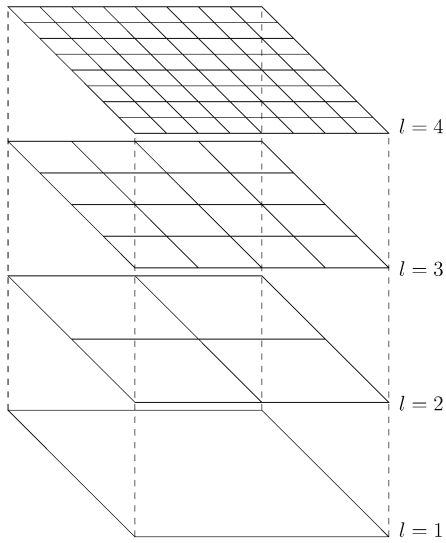

Fig. 3. Partitioning of an image and the resulting four-level hierarchical overlapping block structure.

Given an $N \times M$ sized image X^*, we first form a multilevel hierarchical block structure as shown in Fig. 3. Let us denote a block in this hierarchy by X_l^{ij}, where the indices ij represent the spatial position of the block and l is the level of the hierarchy to which the block belongs. The total number of levels in the hierarchy is further denoted by L.

On the highest level, we partition the image X into size $O \times P$ nonoverlapping blocks $\{X_L^{00}, X_L^{01}, X_L^{10}, \ldots, X_L^{nm}\}$. At each successive (lower) level, the image is partitioned into blocks which, in turn, are composed of 2×2 blocks at the preceding (higher) level of the hierarchy. That is, for $l = L - 1$ to 2

$$\begin{bmatrix} X_{l+1}^{2i,2j} & \| & X_{l+1}^{2i,2j+1} \\ X_{l+1}^{2i+1,2j} & \| & X_{l+1}^{2i+1,2j+1} \end{bmatrix} = X_l^{ij}.$$

Finally, the bottom level of the hierarchy consists of only one block $X_1^{00} = X$. Note that we have larger blocks at lower levels of the hierarchy, with a particular size $2^{L-l}O \times 2^{L-l}P$ at the Lth level. No filtering or decimation is performed.

We embed (and detect) a spread-spectrum watermark for each block at each level of the hierarchy, thereby providing localization capability. The watermarks are embedded using the method of POCS with the constraint sets outlined in Section II-B. The projections onto these sets are described in Appendix B. We note that the blocks at different levels overlap and, therefore, share the total image context provided by the image.

Thus, the watermarks at different levels of the hierarchy contribute interference to each other in addition to the interference encountered from the cover image. In this context, the individual watermark detectability constraints formulated as in (5) ensure that the watermarked image bears each watermark with sufficient strength. The overall image fidelity and point-wise fidelity constraints of (6) and (7) are used to ensure imperceptibility

even under the combined impact of multiple watermarks. Robustness to mild compression for each watermark and fragility under aggressive compression constraints are further ensured by the constraints in (9) and (12), respectively.

IV. Experimental Setup and Results

A. Experimental Setup

We illustrate the efficacy of the set theoretical framework with an implementation of the proposed semifragile watermark for gray-scale images.

1) Algorithm Parameters: The set theoretic framework is a powerful means of simultaneously satisfying multiple constraints in watermarking that often act in opposite directions. The bounds for the constraint sets that determine, for example, acceptable levels of robustness and imperceptibility, would typically be application dependent. As application-based development and extensive testing of these bounds are beyond the scope of this paper, we have empirically determined a set of parameters that provide acceptable levels of visual fidelity, malicious manipulation sensitivity, and robustness/fragility to compression.

In our implementation of the POCS watermarking algorithm of Section II-A, we also terminate at the convergence check step if $K_{\max} = 60$ iterations are exceeded before numerical convergence is achieved. The constraint sets $\{S_i\}_{i=1}^{n}$ in the POCS embedding algorithm described in Section II-A are arranged so that the order of successive projections proceeds in the sequence[4] 1) detectability; 2) fragility under aggressive compression; 3) robustness to compression; 4) overall visual fidelity; 5) pixel-wise image fidelity. While this is irrelevant when the algorithm converges, in scenarios where the algorithm fails to converge, this ordering ensures that pixel-wise visual fidelity is always maintained in the watermarked image. Moreover, the constraints ordered later in the successive projection sequence are empirically favored over the earlier ones. Alternate priorities may dictate a different ordering or relaxation of some constraints.

For the spread-spectrum watermark signal W, we utilize (possibly colored) pseudorandom noise that is obtained by generating white bipolar noise and replicating each sample R times horizontally and vertically in the 2-D image plane. The replication provides a simple method for shaping the spectral characteristics of the watermark to conform better to images that are typically low-pass. Alternate methods for spectrally shaping the watermark in order to make it more compatible with the image, that are motivated by optimal signal design considerations, may also be used instead (see [29] for an example). Note that the compression robustness and imperceptibility constraints already (implicitly) shape the spectrum of the embedded watermark. Nevertheless, choosing a suitable watermark still provides a benefit at the detector, as it ensures that the sequence used for detection is better matched to the watermark actually embedded in the image.

The parameters for the constraint sets used in the algorithm are set as follows: The bound for the overall image fidelity threshold [26] is set at $\theta = 10$ and values of $P_0 = 30$, $P_1 = 3$, $D = 150$ and neighborhood $\nu = 1$ are used for the pixel-wise image fidelity parameters [24]. The embedding strength is set to the linear correlation lowerbound of $\gamma = 1.0 \times MN$, corresponding to a normalized value of $\bar{\gamma} = (\gamma/MN) = 1.0$. In most image blocks, this limit can be satisfied without violating the imperceptibility constraints. The corresponding detection threshold at the receiver is set to $\lambda = 0.6$. The combination of these values provides a good tradeoff between robustness to benign operations (low false alarm rates) and sensitivity to malicious operations (low miss rates). The fragility of the method can be increased (decreased) by increasing (respectively decreasing) the detection threshold.

The quantizer $Q_0[\cdot]$ in (9) is determined by the indices of DCT coefficients quantized to zero at a JPEG quality factor $Q_{\text{limit}} = 60$. This determines the compression resilience. Similarly, the fragility under aggressive compression is achieved by determining the function $Q_0^A[\cdot]$ in (12) by the indices of DCT coefficients quantized to zero at a JPEG quality factor $Q_{\text{limit}}^F = 40$. In order to allow some margin for noise and to take into account the approximation in our definition, the threshold for fragility under aggressive compression is actually set to a normalized value[5] of $\bar{\gamma} = -0.2$. We emphasize that these constraints are approximations (see Section II-B). Thus, meeting these constraints does not fully guarantee that the watermarks survive compression quality factors higher than 60 and vanish for factors lower than 40. Also, the fragility constraint may or may not be required in applications and, therefore, consider cases that include and exclude it.

For hierarchical embedding, we utilize an elementary block-size of 64×64 pixels. This provides a sufficient length for the spread-spectrum sequence (even in the presence of replications) while being small enough to provide good tamper localization capability.

2) Test Images: We used the Goldhill, Lena, Barbara, Mandrill, Washat, Peppers, Boat, and Zelda images from the University of Southern California test image set [31], a hundred miscellaneous Kodak test images, and the Aerial image seen in Fig. 8. These are 8-b gray-scale images with a size of 512×512 pixels for the USC and Kodak images, and 1024×1024 for the Aerial image. Thus, using our 64×64 elementary block size, the USC images and Kodak images have a $L = 4$ level hierarchy for embedding and the Aerial image has a $L = 5$ level hierarchy.

B. Performance

We first examine how we realize advantages of the implicit set-theoretic watermark embedding scheme that were mentioned at the end of Section II-B. For this purpose, in each case, we use a small number of constraint sets that clearly illustrate the benefits realized in the absence of competing requirements. Then, we present the performance of the complete set-theoretic watermarking scheme incorporating all of the constraints together and comment on the tradeoff between conflicting requirements when these cannot be met simultaneously. Note that in order to define a valid watermarking scheme in the set-theoretic framework, visual fidelity and watermark detectability are essential constraints and are therefore required in all cases (in some form).

[4] In (1), this corresponds to the order of increasing index values for the sets.

[5] This value was seen empirically to provide good performance.

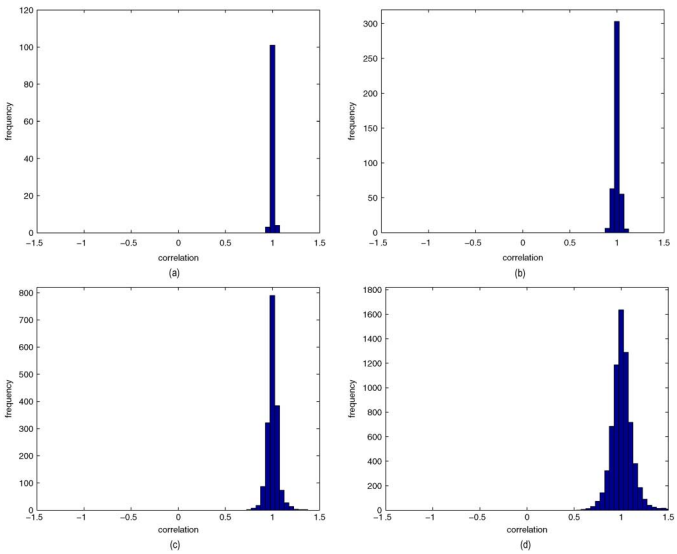

Fig. 4. Histogram of detector responses of 108 different images at various layers. Robustness and fragility sets are excluded during watermark process. (a) Level $l = 1$, (b) Level $l = 2$, (c) Level $l = 3$, (d) Level $l = 4$. (Color version available online at http://ieeexplore.ieee.org.)

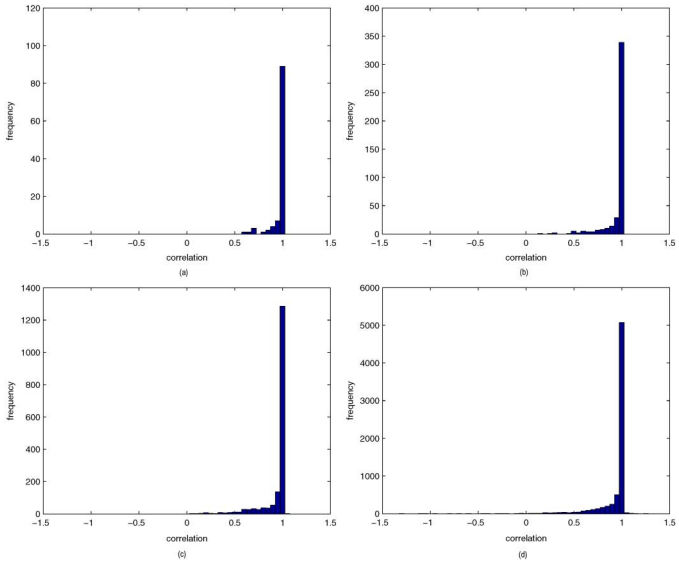

Fig. 5. Histogram of detector responses of 108 different images at various layers. (a) Level $l = 1$, (b) level $l = 2$, (c) level $l = 3$, and (d) level $l = 4$. (Color version available online at http://ieeexplore.ieee.org.)

1) Interference Adaptation: Set theoretic watermarking is an elegant example of informed watermark embedding.[6] From the constraint set definitions, we see that in a feasible image, the interference from the cover image and multiple watermarks at different levels of the hierarchy is completely accounted for in the embedding process. This informed embedding guarantees a uniform embedding strength, when possible and, thus, improved robustness to subsequent processing. We illustrate this property in Fig. 4 which presents histograms of the watermark detector response (mean corrected linear correlations) for all blocks in all 108 test images at different levels of the hierarchy. Constraints used for generating the watermarked images for this case include the primary watermark detectability and visual fidelity constraints. From the histograms in Fig. 4(a)–(d), we see that the detector response is unity or higher for almost all blocks at levels $l = 1, 2$ of the hierarchy (larger blocks). Note that at the lower levels of the hierarchy ($l = 3, 4$), we have a small number of blocks that do not satisfy the sufficient embedding strength constraint. This is because these blocks do not have enough activity to mask the distortion introduced by the imposed watermarks.[7] For these blocks, a sufficient embedding strength is not possible unless corresponding visibility constraints are relaxed at the expense of image fidelity. The histograms indicate that both cover image interference and interwatermark interference across watermarks at different levels of the hierarchy that share the same spatial context are implicitly accounted for in the embedding process.

For the watermarking scenario where all constraint sets are utilized, histograms are shown in Fig. 5. In the presence of the additional constraints, particularly the fragility constraint, the performance degrades slightly and we see that a larger number of blocks fall below the target correlation value of 1.0 at all levels of the hierarchy. For images consisting of mostly smooth regions, the robustness to compression constraint conflicts strongly with the visual fidelity constraint and for these images, even the lowest level ($l = 1$) blocks (consisting of the entire image) fail to meet the target correlation value requirement. Overall, however, the histograms illustrate that the method does an excellent job of maintaining watermark strength in most blocks despite the contrary constraints.

2) Visual Fidelity Adaptation: The impact of the visual fidelity constraints of the hierarchical watermarking scheme is illustrated on Goldhill image (Fig. 6). Fig. 6(a) shows the original image and Fig. 6(b) illustrates the watermarked image obtained with the hierarchical semifragile watermarking scheme using the parameters as described earlier and utilizing all constraint sets. The watermarked image shows minimal artifacts despite the fact that each pixel carries the payload for the four watermarks at the different levels of the hierarchy. The imperceptibility constraints imposed on the watermark effectively prevent significant perceptible artifacts. The only noticeable artifact appears at the upper part of the image in the sky region that is extremely smooth and, therefore, does not allow embedding without concurrent distortion. To specifically illustrate the benefit of the visual fidelity constraints, we compare the performance without this constraint but with an overall peak signal-to-noise ratio (PSNR) constraint that is selected to match the PSNR of Fig. 6(b). This is obtained within our framework by dropping the pixel-wise image fidelity constraint entirely and replacing the operator H in the overall visual fidelity constraint in (6) with the identity operator so that it defines a bound on the PSNR, which is defined as $\mathrm{PSNR} = 10\log_{10}((255^2)/\mathrm{MSE})$. Note that this also illustrates the flexibility of the set theoretic embedding approach: the nature of the embedding can be changed from a visually adaptive method to one that only has a total-power constraint by a simple redefinition of a set. As seen in Fig. 6(c), without the visual fidelity constraint, the image has more artifacts (at the same overall PSNR) which can be particularly seen in the sky region. A closer view of this region is shown

[6]Recall that we use the term informed-embedding as defined in [22, p. 132], which is distinct from side-informed coding methods.

[7]Particularly, the Kodak image database [32] includes several images with extremely large smooth regions.

Fig. 6. Goldhill image illustrating the impact of visual fidelity constraint. $\mathrm{Size} = 512 \times 512$, $L = 4$, $R = 2$. (a) Original. (b) Watermarked with visual fidelity constraint ($\mathrm{PSNR} = 31.87$). (c) Watermarked without visual fidelity constraint ($\mathrm{PSNR} = 32.37$ dB). (d) Watermarked without robustness and fragility constraints ($\mathrm{PSNR} = 40.09$ dB).

in Section IV-B2 in Fig. 7 where the advantage of the visual fidelity constraint can be clearly seen.[8]

3) Compression Resilience: In this section, we illustrate the resilience of the proposed watermark to compression and the impact of the robustness to the compression set. For the results in this section, we exclude the fragility under aggressive compression constraint and defer results and discussion for that constraint to the next section. Table I summarizes the results for watermark detection for the proposed scheme across the 108 images from the USC and Kodak image sets for JPEG compression with Q-factors ranging from 90 down through 10.

The impact of shaping the watermark spectral characteristics through the use of the replication factor R (by which the watermark pseudonoise sequence is replicated along each of the two image dimensions, the value seen in the leftmost column) is also included within this table. Each entry in the table is represented as a/b where b is the total number of blocks at that level of the hierarchy encountered in the experiment and a is the number of these blocks in which the watermark is successfully detected (for the corresponding compression level). The performance of the detection in the absence of the robustness to compression constraint is summarized for the eight images in the USC data set in Table II for comparison.

From the data in the tables, several observations can be made. A comparison of the $R = 1$ (i.e., a spectrally white watermark) entries in Tables I and II illustrates the significant impact of the robustness to compression constraint. With the constraint, the watermark is detected for a large majority of the blocks for JPEG quality factors above 50, whereas the detection performance degrades very severely without the constraint. The watermark is

[8]In this example, the spatial localization constraint and hierarchical embedding limit the capability of the visual fidelity constraint to redistribute watermark power. In scenarios where the method is utilized to embed multiple bits over the entire image, this constraint is absent and the effect is more significant [23].

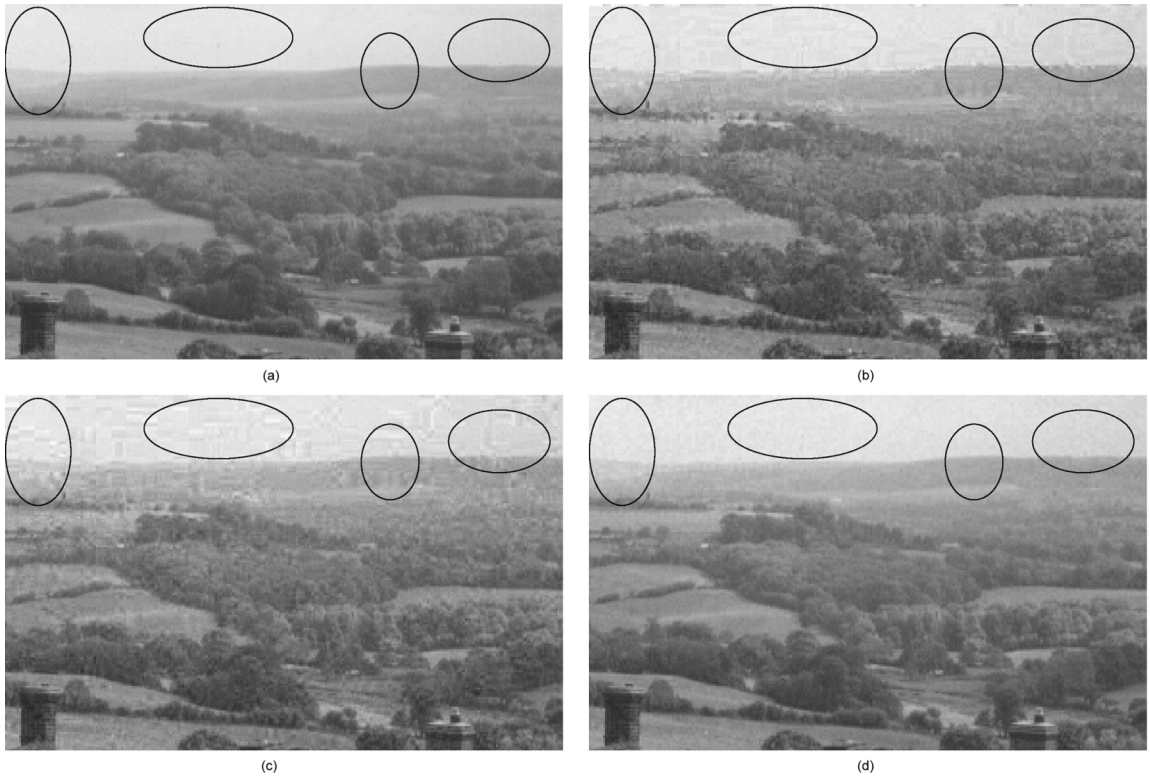

Fig. 7. Closeup of a region of the Goldhill image illustrating the impact of visual fidelity, robustness, and fragility constraints. Regions that are significantly different are encircled. (a) Original Goldhill image. (b) Watermarked with visual constraint. (c) Watermarked without visual constraint. (d) Watermarked without robustness and fragility constraint.

TABLE I
SUMMARY OF WATERMARK DETECTION RESULTS FOR ~~EIGHT~~ WATERMARKED IMAGES AT VARIOUS JPEG COMPRESSION LEVELS. THE ROBUSTNESS TO COMPRESSION IS SET AT JPEG QUALITY FACTOR 60 AND FRAGILITY SET IS EXCLUDED

Repl. R	Level l	$Q = 90$	$Q = 80$	$Q = 70$	$Q = 60$	$Q = 50$	$Q = 40$	$Q = 30$	$Q = 20$	$Q = 10$
1	1	108/108	108/108	108/108	108/108	107/108	105/108	101/108	91/108	0/108
	2	423/432	421/432	420/432	413/432	412/432	403/432	386/432	347/432	59/32
	3	1676/1728	1662/1728	1622/1728	1565/1728	1534/1728	1490/1728	1413/1728	1225/1728	415/1728
	4	6650/6912	6566/6912	6290/6912	6013/6912	5766/6912	5492/6912	5182/6912	4222/6912	1967/6912
2	1	108/108	108/108	108/108	108/108	108/108	108/108	108/108	108/108	104/108
	2	431/432	431/432	431/432	431/432	431/432	431/432	431/432	431/432	368/432
	3	1699/1728	1698/1728	1698/1728	1697/1728	1695/1728	1695/1728	1691/1728	1676/1728	1290/1728
	4	6737/6912	6728/6912	6704/6912	6711/6912	6688/6912	6663/6912	6629/6912	6449/6912	4308/6912

TABLE II
WATERMARK DETECTION RESULTS IN ABSENCE OF ROBUSTNESS TO COMPRESSION SET AT VARIOUS JPEG COMPRESSION LEVELS

Repl. R	Level l	$Q = 90$	$Q = 80$	$Q = 70$	$Q = 60$
1	1	7/8	1/8	0/8	0/8
	2	26/32	6/32	4/32	2/32
	3	85/128	24/128	15/128	13/128
	4	337/512	162/512	128/512	106/512

detected only in a minority of the blocks even at $\mathrm{Q} = 80$. By comparing the entries for $R = 1$ and $R = 2$ in Table I, we also observe that spectrally shaping the watermark by replication further improves robustness to compression. As we indicated earlier in this section and at the end of Section II-B1, this improvement comes from the detector. The choice of the watermark sequence is not constrained by the set-theoretic framework.

From Table I, we also see that the detection performance improves at lower levels of the hierarchy (small l, larger blocks). This is due to the lengthening of the spreading sequence with the increased block size. The process is beneficial since greater robustness is often desirable for the lower levels of the hierarchy. The detection performance at different levels may also be controlled by establishing different embedding strength thresholds and different robustness to compression bounds. Also note that

TABLE III
SUMMARY OF WATERMARK DETECTION RESULTS FOR EIGHT WATERMARKED IMAGES AT VARIOUS JPEG COMPRESSION LEVELS. THE ROBUSTNESS TO COMPRESSION IS SET AT JPEG QUALITY LEVEL 60 AND FRAGILITY TO AGGRESSIVE COMPRESSION SET USING AT JPEG QUALITY LEVEL 40

Repl. R	Level l	$Q = 90$	$Q = 80$	$Q = 70$	$Q = 60$	$Q = 50$	$Q = 40$	$Q = 30$	$Q = 20$	$Q = 10$
1	1	107/108	105/108	102/108	98/108	84/108	77/108	64/108	5/108	0/108
	2	413/432	407/432	396/432	367/432	317/432	292/432	239/432	39/432	0/432
	3	1621/1728	1574/1728	1470/1728	1339/1728	1095/1728	965/1728	843/1728	282/1728	27/1728
	4	6526/6912	6140/6912	5578/6912	4768/6912	3914/6912	3486/6912	3064/6912	1769/6912	540/6912
2	1	107/108	103/108	99/108	97/108	94/108	91/108	87/108	74/108	2/108
	2	408/432	411/432	405/432	429/432	405/432	387/432	361/432	329/432	43/432
	3	1608/1728	1689/1728	1551/1728	1499/1728	1436/1728	1373/1728	1254/1728	1109/1728	391/1728
	4	6457/6912	6343/6912	6245/6912	5743/6912	5267/6912	4922/6912	4328/6912	3704/6912	1982/6912

when $R = 1$, for a few blocks at $l = 4$ (smallest blocks) and for one block at $l = 3$, the watermark is not detectable even at Q $= 90$. As mentioned in Section IV-B1, these blocks correspond to extremely smooth regions where no embedding is possible without violating the visual fidelity constraint.[9]

4) Aggressive Compression Fragility: In this section, we consider the impact of the constraint imposing fragility under aggressive compression. In order to do so, we incorporate all constraints in our POCS watermarking scheme and examine the performance in the presence of compression as was done in the previous section. Table III lists the results for watermark detection under this operational scenario for a number of compression ratios in the same format as Table I. From Table III, we see that the watermarks start to deteriorate under vigorous compression starting $Q = 40$ at all levels of the hierarchy. The efficacy of the fragility set is clear at $Q = 20$ and below. This is particularly clear when we compare the results in Table III against those in Table I, where only the robustness to compression set was used.

Observe that the robustness to compression and fragility under robust compression constraints act in diametrically opposite directions and, therefore, are the most difficult to satisfy concurrently. This conflict is also apparent from a comparison of Tables I and III. In interpreting the results for Table III, we would also like to recall the discussion from Section IV-A1 regarding the implicit prioritization of constraints due to their ordering. As indicated, our ordering favors robustness to compression over fragility under aggressive compression. This (and the approximate nature of the set) explains why the watermarks are not eliminated precisely at the JPEG quality factor $Q^F_{\text{limit}} = 40$ used for the fragility constraint. Experimental validation with reordering the sets indicates that indeed the fragility may be given priority over robustness to compression using this process.

5) Tamper Detection Capability: Tamper detection capability of the scheme is illustrated with a malicious manipulation example. A five-level hierarchical watermark is embedded within the Aerial image for which the original is shown in Fig. 8(a). The watermarked image shown in Fig. 8(b) is obtained by the proposed POCS algorithm using all sets except fragility and with robustness to JPEG compression determined by the quality factor $Q_{\text{limit}} = 30$ and other parameters as specified before. After embedding of the watermark, the image is compressed using a JPEG quality factor of 50. The resulting watermarked and compressed image is then tampered with to remove vehicles around the center building. This tampered version of the image is shown in Fig. 8(c). The results obtained from the watermark detection process applied to the watermarked-compressed-tampered image are shown in Fig. 8(d). Image regions for which the watermark detection fails at any hierarchy level are indicated by shading and the corresponding hierarchy-level number at which the watermark detection (first) fails is overlaid on these shaded regions. The detection results correctly identify the tampered regions and localize the tampering.

6) Computational Complexity: For the parameters in Section IV-A1 and the eight images from the USC database with a four-level hierarchy, the POCS iterations for the watermark embedding converge to a feasible solution in typically 15 iterations. This takes approximately 4 min on a Pentium M 1.70-GHz machine with Matlab implementation. In instances where the algorithm does not converge after 60 iterations (approximately 15 min), the process is terminated and the current iterate is used as the approximate solution.

The proposed framework constitutes a middle ground between heuristic embedding methods, which can guarantee neither robustness nor imperceptibility, and optimization-based methods, which require substantial computational resources. Moreover, incorporating linear and nonlinear constraints in the system is significantly easier compared with the optimization methods.

As a result, the algorithm provides a powerful practical alternative to both heuristic and optimization-based methods, especially in commercial content distribution applications where preserving the perceptual quality is paramount.

V. CONCLUSION

In this paper, we introduced a set theoretic framework for watermarking that formulates watermark embedding as a problem of determining a feasible solution meeting multiple constraints. We have defined constraint sets for common watermarking requirements of detectability, robustness against compression, imperceptibility, and fragility under aggressive compression and have developed an algorithm for watermark embedding using the method of POCS. We implemented a hierarchical semifragile watermark in the framework and illustrated

[9]This is not a severe problem as one may also question the need for integrity verification of constant blocks in isolation without their surrounding context.

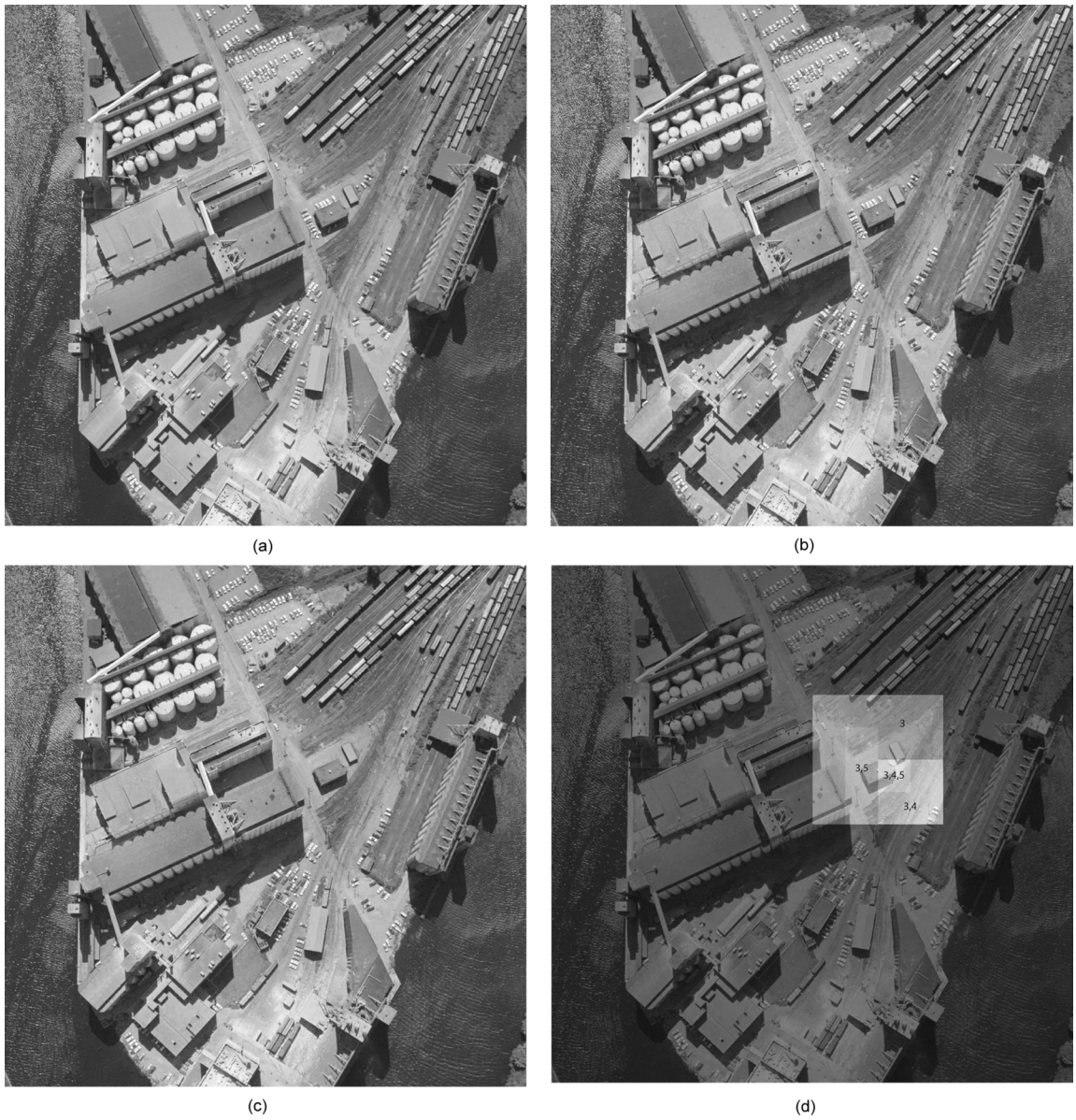

Fig. 8. Tamper localization performance on the aerial image. Size 1024×1024 pixels. Image in (c) has been manipulated by removing the vehicles around the building at the center. Detection results on the manipulated image after JPEG compression with $Q = 50$ are shown in (d). Numbers indicate the levels at which the watermark is not detected. (a) Original image. (b) Watermarked image ($L = 5$, $R = 2$, $\mathrm{PSNR} = 36.61$ dB). (c) Manipulated image. (d) Detection result ($\mathrm{Q} = 50$).

the power of the framework. Specifically, the watermarking scheme automatically and implicitly handles both host signal interference and interwatermark interference, ensures visual fidelity, provides robustness against mild JPEG compression, and fragility under aggressive JPEG compression and the capability to detect alterations under mild compression. The set theoretic watermarking framework successfully unifies multiple watermarking, visually adaptive watermarking, and interference cancellation scenarios.

Appendix A
Texture Masking Model

We adopt the spatial texture masking model developed in [34] that was previously used for watermark embedding in [24]. We present a brief overview of the model here and refer the reader to these publications for greater detail. The cover image is modeled as a sum of the local mean and an error term, with the latter further modeled by a generalized Gaussian distribution. A noise visibility function (NVF) at each pixel position (i, j) is obtained from this model as

$$\mathrm{NVF}(i,j) = \frac{U(i,j)}{U(i,j) + \sigma_X^2} \tag{13}$$

where

$$U(i,j) = \gamma[\eta(\gamma)]^{\gamma} \frac{1}{||R(i,j)||^{2-\gamma}}; \quad \eta(\gamma) = \frac{\Gamma\left(\frac{3}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma}\right)} \tag{14}$$

$$R(i,j) = \frac{X(i,j) - \bar{X}(i,j)}{\sigma_X}; \quad \Gamma(t) = \int_0^{\infty} e^{-v} v^{(t-1)}\, dv. \tag{15}$$

The shape parameter γ and local image variance σ_X^2 specify the model. The shape parameter is image dependent and estimated by moment matching. It ranges between 0.3 and 1.0 for most real images.

The maximum likelihood estimator of the variance is utilized, which is given by

$$\sigma_X^2(i,j) = \frac{1}{(2\nu+1)^2} \sum_{k=-\nu}^{\nu} \sum_{l=-\nu}^{\nu} (X(i+k,j+l) - \bar{X}(i,j))^2 \tag{16}$$

where ν defined the neighborhood size over which the image mean is computed in the model and

$$\bar{X}(i,j)) = \frac{1}{(2\nu+1)^2} \sum_{k=-\nu}^{\nu} \sum_{l=-\nu}^{\nu} X(i+k,j+l).$$

We use a simplified empirical version of the NVF model [33]

$$\mathrm{NVF}(i,j) = \frac{1}{1+\kappa\sigma_X^2(i,j)} \tag{17}$$

where tuning parameter κ controls the contrast adjustment $\kappa = (D/(\sigma_{X\,\max}^2))$ where $\sigma_{X\,\max}^2$ is the maximal local variance for a given image and D is an experimentally determined parameter [33], [34].

Using the NVF, allowable pixel distortions are computed as

$$\Delta(i,j) = (1 - \mathrm{NVF}(i,j))P_0 + \mathrm{NVF}(i,j)P_1 \tag{18}$$

where P_0 and P_1 represent the allowable pixel distortion in busy and flat regions, respectively. We use a neighborhood size of $\nu = 1$, and values of $P_0 = 30$ and $P_1 = 3$, based on [24]. The upper and lower bounds on distortion are then set equal to Δ (i.e., $D_U = D_L = \Delta$ is utilized in (7)). In "flat" image regions, the NVF is close to 0 which makes the $(1 - \mathrm{NVF}(i,j))\,P_0$ term small and, therefore, the allowable distortion level $\Delta(i,j)$ is also small. This is consistent with the fact that noise is less visible in busy regions and more visible in flat regions.

Appendix B
Constraint Set Projection Operators

The projection onto each constraint set is computed as follows.

A. Strength of Spread-Spectrum Embedding

Projection of Y onto S_1 is given by

$$P_1(Y) = \arg\min_Z \|Z - Y\|^2$$

subject to

$$W^T(Z - \bar{Z}) \geq \gamma.$$

The Lagrangian [30] for this constrained optimization problem can be written as

$$\mathcal{L} = \|Z - Y\|^2 + \lambda(W^T(T_L Z - \overline{T_L Z}) - \gamma) \tag{19}$$

where T_L is the matrix representing the linear transformation $\mathcal{T}_{\mathcal{L}}$.

The Lagrange parameter is readily shown to be

$$\lambda = \begin{cases} \frac{2[\gamma - W^T T_L(Y-\bar{Y})]}{W^T T_L \left(T_L^T W - \overline{T_L^T W}\right)}, & \gamma > W^T T_L(Y-\bar{Y}) \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

The projection can then be expressed in terms of the Lagrange parameter as

$$P_1(Y) = Y + \frac{\lambda}{2}\left[T_L^T W - \overline{T_L^T W}\right]. \tag{21}$$

B. Projection Onto Overall Fidelity

The projection of Y onto S_2 is given by

$$P_2(Y) = \arg\min_Z \|Z - Y\|^2$$

subject to

$$\|H(Z - X_0)\|^2 \leq \theta^2.$$

Using the method of Lagrange multipliers, the projection is determined as

$$P_1(Y) = X_0 + (I + \lambda H^T H)(Y - X_0) \tag{22}$$

where I denotes the identity matrix, and λ is the Lagrange parameter. If $Y \notin S_2$, the Lagrange parameter is the positive root of

$$(Y - X_0)^T (I + \lambda H^T H)^{-1} H^T H (I + \lambda H^T H)^{-1} (Y - X_0) = \theta^2 \tag{23}$$

and λ is zero if $Y \in S_2$.

If H is assumed to be shift-invariant, as in the case in this paper, the computation of the Lagrange parameter can be significantly simplified by using the discrete Fourier transform (DFT). For $Y \notin S_2$, the Lagrange parameter is the positive root of

$$\frac{1}{NM} \sum_{l,k} \frac{|H(l,k)(Y(l,k) - X_0(l,k))|^2}{(1 + \lambda |H(l,k)|^2)^2} = \theta^2 \tag{24}$$

where $H(l,k), Y(l,k)$, and $X_0(l,k)$ represent the 2-D DFT coefficients of H, Y, and X_0, respectively.

C. Projection Onto Point-Wise Fidelity

The projections onto the point-wise fidelity constraint set are readily determined term-wise. If we denote $V = P_3(Y)$ then for each $0 \leq j < (MN-1)$, the element V_j is given by

$$V_j = \begin{cases} U_j, & \text{if } Y_j > U_j \\ L_j, & \text{if } Y_j < L_j \\ Y_j, & \text{otherwise} \end{cases} \tag{25}$$

where U_j, L_j, and Y_j are the jth elements of U,, L, and Y, respectively.

D. Projection Onto Robustness to Compression

Note that the robustness to compression set in (26) can be rewritten as

$$\hat{S}_4 \equiv \{X : W^T(\mathcal{T}_{\mathcal{C}}X - \overline{\mathcal{T}_{\mathcal{C}}X}) \geq \gamma\} \tag{26}$$

where $\mathcal{T}_{\mathcal{C}} = \mathcal{T}_{\mathcal{I}} Q_0 \mathcal{T}_{\mathcal{F}}$ is the concatenation of the linear operators $\mathcal{T}_{\mathcal{I}}$, Q_0, and $\mathcal{T}_{\mathcal{F}}$. Now the set has the same form as the set in (5) and we can use the projection derived earlier for this set. It is worth noting that the operator $\mathcal{T}_{\mathcal{C}}$ is normally singular.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their detailed comments that have significantly improved this paper. Particularly, the explicit modeling of fragility under aggressive compression in their framework was prompted by reviewer suggestions.

REFERENCES

[1] A. Menezes, P. van Oorschot, and S. Vanstone, *Handbook of Applied Cryptography*. Boca Raton, FL: CRC, 1997.
[2] A. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 758–767, Feb. 2005.
[3] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proc. IEEE*, vol. 86, no. 6, pp. 1064–1087, Jun. 1998.
[4] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul. 1999.
[5] I. J. Cox, J. Killian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.
[6] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.
[7] M. Yeung and F. Mintzer, "Invisible watermarking for image verification," *J. Electron. Imag.*, vol. 7, no. 6, pp. 578–591, Jul. 1998.
[8] P. W. Wong, "A public key watermark for image verification and authentication," in *Proc. IEEE Int. Conf. Image Processing*, Chicago, IL, Oct. 4–7, 1998, pp. 425–429.
[9] J. Fridrich, M. Goljan, and A. C. Baldoza, "New fragile authentication watermark for images," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, BC, Canada, Sep. 10–13, 2000, pp. 446–449.
[10] D. Coppersmith, F. C. Mintzer, C. P. Tresser, C. W. Wu, and M. M. Yeung, "Fragile imperceptible digital watermark with privacy control," in *Proc. SPIE Security and Watermarking of Multimedia Contents I*, Jan. 1999, vol. 3657, pp. 79–84.
[11] M. U. Celik, G. Sharma, E. Saber, and A. M. Tekalp, "Hierarchical watermarking for secure image authentication with localization," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 585–595, Jun. 2002.
[12] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proc. IEEE*, vol. 87, no. 7, pp. 1167–1180, Jul. 1999.
[13] J. Eggers and B. Girod, "Blind watermarking applied to image authentication," in *Proc. IEEE ICASSP*, Salt Lake City, UT, May 2001, pp. 1977–1980.
[14] S. Bhattacharjee and M. Kutter, "Compression tolerant image authentication," in *Proc. IEEE Int. Conf. Image Processing*, Chicago, IL, Oct. 1998, pp. 435–439.
[15] L. M. Bregman, "The method of successive projection for finding a common point of convex sets," *Dokl. Akad. Nauk. SSSR*, vol. 162, no. 3, pp. 688–692, 1965.
[16] H. Stark and Y. Yang, *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics*. New York: Wiley, 1998.
[17] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, no. 2, pp. 182–208, Feb. 1993.
[18] K. C. Haddad, H. Stark, and N. P. Galatsanos, "Constrained FIR filter design by the method of vector space projections," *IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process.*, vol. 47, no. 8, pp. 714–725, Aug. 2000.
[19] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 2, pp. 201–212, Apr. 1984.
[20] G. Sharma and H. J. Trussell, "Set theoretic signal restoration using an error in variables criterion," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1692–1697, Dec. 1997.
[21] E. J. Delp, C. I. Podilchuk, and E. T. Lin, "Detection of image alterations using semi-fragile watermarks," in *Proc. SPIE Int. Conf. Security and Watermarking of Multimedia Contents II*, Jan. 2000, vol. 3971, no. 5, pp. 409–413.
[22] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. San Mateo, CA: Morgan Kaufmann, 2001.
[23] O. Altun, G. Sharma, and M. Bocko, "Informed watermark embedding in fractional Fourier domains," presented at the EUSIPCO, Antalya, Turkey, Sep. 2005, unpublished.
[24] S. Pereira, S. Voloshynoskiy, and T. Pun, "Optimal transform domain watermark embedding via linear programming," *Signal Process.*, vol. 81, no. 6, pp. 1251–1260, Jun. 2001.
[25] G. Sharma, "Color fundamentals for digital imaging," in *Digital Color Imaging Handbook*. Boca Raton, FL: CRC, 2003, ch. 1.
[26] J. L. Mannos and D. L. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 525–536, Jul. 1974.
[27] A. B. Watson, *Digital Images and Human Vision*. Cambridge, MA: MIT Press, 1993.
[28] P. Moulin, "Information-Hiding Games," *Springer-Verlag Lect. Notes Comput. Sci.*, vol. 2613, pp. 1–12, 2002.
[29] J. K. Su, J. J. Eggers, and B. Girod, "Analysis of digital watermarks subjected to optimum linear filtering and additive noise," *Signal Process., Special Issue Inf. Theoretic Issues Digital Watermarking*, vol. 81, no. 6, pp. 1141–1175, 2001.
[30] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. Reading, MA: Addison-Wesley, 1989.
[31] [Online]. Available: http://sipi.usc.edu/database/.
[32] *Kodak PhotoCD Images*, , Eastman Kodak Company [Online]. Available: ftp://ftp.kodak.com/www/images/pcd.
[33] S. Efstratiadis and A. Katsaggelos, "Adaptive image restoration with reduced computational load," *Opt. Eng.*, vol. 29, no. 12, pp. 845–853, Dec. 1990.
[34] S. Voloshynovskiy, A. Herrigel, N. Baumgeartner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," in *Proc. 3rd Int. Workshop Information Hiding*, Sep. 29–Oct. 1, 1999, pp. 211–236.

Oktay Altun (S'03) received the B.Sc. degree in electrical and electronic engineering from Bilkent University, Ankara, Turkey, in 2000 and the M.Sc. degree in electrical and computer engineering from the University of Rochester, Rochester, NY, in 2005.

Currently, he is a Research Assistant with the Electrical and Computer Engineering Department, University of Rochester. His research interests include multimedia security, digital watermarking, image processing, and analog circuit optimization.

Gaurav Sharma (SM'00) received the B.E. degree in electronics and communication engineering from the Indian Institute of Technology Roorkee, Roorkee, India, in 1990; the M.E. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1992; and the M.S. degree in applied mathematics and the Ph.D. degree in electrical and computer engineering from North Carolina State University (NCSU), Raleigh, in 1995 and 1996, respectively.

From 1992 to 1996, he was a Research Assistant with the Center for Advanced Computing and Communications in the Electrical and Computer Engineering Department at NCSU. From 1996 to 2003, he was with Xerox Research and Technology, Webster, NY, as a Member of the

Research Staff and subsequently as Principal Scientist. Currently he is an Associate Professor in the Department of Electrical and Computer Engineering and in the Department of Biostatistics and Computational Biology at the University of Rochester, Rochester, NY. His research interests include multimedia security and watermarking, color science and imaging, and signal processing for visual sensor networks.

Dr. Sharma is the Editor of the *Color Imaging Handbook* (CRC, 2003). He is a member of Sigma Xi, Phi Kappa Phi, Pi Mu Epsilon, and IS&T. He was the 2003 Chair for the Rochester chapter of the IEEE signal processing society and is the Treasurer for the Rochester section. He is a member of the IEEE Signal Processing Society's Image and multidimensional signal processing (IMDSP) technical committee, and the Standing committee on Industry DSP. He is an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and the *Journal of Electronic Imaging*.

Mehmet U. Celik (S'98–M'06) received the B.Sc. degree in electrical and electronic engineering from Bilkent University, Ankara, Turkey, in 1999 and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Rochester, Rochester, NY, in 2001 and 2004, respectively.

From 1999 to 2004, he was a Research Assistant with the Electrical and Computer Engineering Department, University of Rochester. In 2001 and 2002, he was a Research and Development Intern with Digimarc Corporation, Tualatin, OR; and in 2003 with Sharp Labs of America, Camas, WA. He has been with the Information and Systems Security Department, Philips Research, Eindhoven, The Netherlands, since 2005. His research interests include multimedia security, digital watermarking, image and video processing, and cryptography.

Mark F. Bocko (M'94) received the B.A. degree in physics and astronomy from Colgate University, Hamilton, NY, in 1978 and the Ph.D. degree in physics from the University of Rochester, Rochester, NY, in 1984.

Currently, he is Full Professor and Chairman of the Department of the Electrical and Computer Engineering Department, University of Rochester. His research interests span a number of areas, including sensors and integrated sensor systems, audio and music signal processing, precision measurements, superconducting electronics quantum noise, and quantum computing.