Spatial Domain Digital Watermarking of Multimedia Objects for Buyer Authentication

Dipti Prasad Mukherjee, *Member, IEEE*, Subhamoy Maitra, and Scott T. Acton, *Senior Member, IEEE*

***Abstract*—Most of the existing watermarking processes become vulnerable when the attacker knows the watermark insertion algorithm. This paper presents an invisible spatial domain watermark insertion algorithm for which we show that the watermark can be recovered, even if the attacker tries to manipulate the watermark with the knowledge of the watermarking process. The process incorporates buyer specific watermarks within a single multimedia object, and the same multimedia object has different watermarks that differ from owner to owner. Therefore recovery of this watermark not only authenticates the particular owner of the multimedia object but also could be used to identify the buyer involved in the forging process. This is achieved after spatially dividing the multimedia signal randomly into a set of disjoint subsets (referred to as the *image key*) and then manipulating the intensity of these subsets differently depending on a buyer specific key. These buyer specific keys are generated using a secret permutation of error correcting codes so that exact keys are not known even with the knowledge of the error correcting scheme. During recovery process a manipulated buyer key (due to attack) is extracted from the knowledge of the image key. The recovered buyer key is matched with the exact buyer key in the database utilizing the principles of error correction. The survival of the watermark is demonstrated for a wide range of transformations and forging attempts on multimedia objects both in spatial and frequency domains. We have shown that quantitatively our watermarking survives rewatermarking attack using the knowledge of the watermarking process more efficiently compared to a spread spectrum based technique. The efficacy of the process increases in scenarios in which there exist fewer numbers of buyer keys for a specific multimedia object. We have also shown that a minor variation of the watermark insertion process can survive a "Stirmark" attack. By making the image key and the intensity manipulation process specific for a buyer and with proper selection of error correcting codes, certain categories of collusion attacks can also be precluded.**

***Index Terms*—Buyer key, digital watermarking, error correcting code, image key.**

I. Introduction

In THIS paper, we present an invisible digital watermarking technique for multimedia objects. The watermark that we are introducing in the multimedia object is in the form of a bit pattern specific for an individual buyer. We show that effective recovery of this bit pattern is possible under a variety of nontrivial attacks.

In our approach, we assume a possible forger knows the proposed watermarking algorithm. So, we are specifically investigating the process of watermarking when the encoding algorithm is known. Given this assumption, most of the existing watermarking techniques are vulnerable. The question of copyright protection should ideally accommodate for both intentional attacks and common image transformations, such as rotation, scaling, filtering *etc*., on the watermarked image [11], [14]. Ruanaidh *et al.* [15] have assessed an exhaustive list of possible threats and exploitation for digital watermarking in images.

Voloshynovskyi *et al.* [16] have introduced a set of attacks in a so-called s*econd generation watermarking benchmark*. They have included four categories of attacks, namely removal, geometric, cryptographic and protocol attacks. We will address them individually while presenting our simulation results. We note that the cryptographic category of attacks mentioned in [16] should include a specific attack possible due to the knowledge of watermarking scheme, which is the central focus of this paper.

Important watermark insertion strategies revolve around inserting watermarks in the *perceptually significant* regions of the image [9]. This motivation is based on the fact that any attempt to modify the watermark results in visible distortion of the image. Cox *et al.* have proposed spread spectrum based insertion of watermarks by manipulating the discrete cosine transform (DCT) components [2]. The inserted watermark is recovered using a statistical similarity measure with the original watermark. A similar approach using statistical modeling of the DCT coefficients is reported in [7]. The differential-energy watermarking algorithm embeds labeled bits by selectively discarding high frequency DCT coefficients in certain selective image regions [10]. For all these schemes, knowledge of watermarking algorithm weakens, if not defeats, the robustness of the process.

Our proposed technique of watermarking does not depend on the perceptually significant regions of the image; rather it is based on the concept of utilizing an image key and a buyer key. The buyer key is a binary bit pattern. It ensures a footprint specific to the buyer of a particular multimedia object. The image key is dependent on the spatial organizations of pixels. The recovery of a watermark in this case not only protects the copyright but also authenticates the possible owner in case multiple copies of the same image or some modifications of it are sold. This authentication process is achieved without any additional computational cost to the watermarking process. Moreover, the process does not degrade the quality of the signal.

Manuscript received November 30, 20012; revised July 30, 2002. The work of S. T. Acton was supported in part by the National Science Foundation under Grant DUE 01211596. The associate editor coordinating the review of this paper and approving it for publication was Dr. Hong Heather Yu.

D. P. Mukherjee is with the Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta, India 700108 (e-mail: dipti@isical.ac.in).

S. Maitra is with the Computer and Statistical Service Center, Indian Statistical Institute, Calcutta, India 700108 (e-mail: subho@isical.ac.in).

S. T. Acton is with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22094 USA (e-mail: acton@virginia.edu).

Digital Object Identifier 10.1109/TMM.2003.819759

Several watermarking techniques exist that introduce watermarks in the spatial domain [3], [4], [17]. In a similar context, an information theoretic model for steganography is given in [1] where uncertainty about the embedded watermark is resolved using principles of hypothesis testing. Most of these techniques do not survive intentional attacks in the frequency domain. In our proposed approach, we show that watermarking can survive a variety of forging attempts involving manipulations in the frequency domain.

Honsinger *et al.* [8] have introduced the iconic message as watermark after dividing the image into a set of disjoint sequential blocks. The watermark is a flat spectrum random phase carrier convolved with an iconic message introduced in each of the image blocks. The quality of the inserted iconic message degrades severely in the case where the image is rewatermarked using a different flat spectrum random phase carrier signal convolved with another iconic message. In our approach, the population of individual image block is randomly selected from different spatial locations of the image matrix. While the recovery in [8] requires a threshold of the correlation between phase carrier and information from the data embedded message, our process does not need any thresholding mechanism, making the watermark recovery process exact.

We have used the principles of error-correcting codes to recover the buyer key that authenticates the ownership of a multimedia object. The proposed watermark insertion and retrieval techniques do not depend on the specific error-correcting algorithm. We assume that the buyer and the image key associated with the process are secure and they are neither image dependent nor algorithm dependent. The process of watermark insertion is controlled to the extent that the image intensity variation does not lead to noticeable distortion.

We focus on the watermarking process as implemented on a two-dimensional image. Identical watermarking can be achieved on one-dimensional audio signals and on video sequences. In Section II, we present the generation of cryptographic parameters, *viz.*, the image and the buyer keys. In Section III, the proposed watermarking scheme is outlined including the watermark recovery process. The number of buyer keys depends on the number of copies disseminated. So, parameters for key generation can be different for an expensive multimedia object compared to a low cost object. Here, it is presumed that the high value items are sold less frequently. Accordingly, a high value item is secured in greater detail compared to a low value one. This is also explained in Section III. The simulations for watermarking, attack and buyer key retrieval results are presented in Section IV. The theory that allows survival of collusion attacks is presented in Section V, along with associated results. This is followed by the conclusion in Section VI.

II. Generation of Image and Buyer Key

In this section we define both the image and the buyer keys including their cryptographic properties to be used for the watermarking purpose.

	0	1	2	3
0	5	7	11	233
1	34	14	79	61
2	123	211	47	*11*
3	0	254	1	191

Fig. 1. Example matrix for the image I_e.

	0	1	2	3
0	3	0	3	3
1	2	0	1	1
2	2	1	2	0
3	1	2	0	3

Fig. 2. Label matrix for the image I_e.

A. Image Key

Consider an image I, which is a matrix of size $2^a \times 2^b$. Let us consider that the image is divided into $m = 2^n$ subgroups, each containing 2^{a+b-n} pixel locations. Let us denote the subgroups by $G_0, G_1, \ldots, G_{m-2}, G_{m-1}$. Each subgroup G_k is thus a set of 2^{a+b-n} tuples of the form $\langle$ *row index, column index* $\rangle$. Note that row index i varies from 0 to $2^a - 1$ and the column index j varies from 0 to $2^b - 1$. It is also clear that $G_{k_1} \cap G_{k_2} = \phi$ for $0 \leq k_1 \neq k_2 \leq m - 1$.

Consider the $2^2 \times 2^2$ matrix in Fig. 1. This matrix corresponds to an example image I_e with $a = b = 2$. Each location of the matrix can be referred as $\langle i, j \rangle$ pair for $0 \leq i \leq 2^a - 1, 0 \leq j \leq 2^b - 1$. Each location $\langle i, j \rangle$ contains some value, typically within 0 to 255 for an 8-bit intensity image. At this point, we take $m = 2^n = 2^2 = 4$ for the example, and we have:

$$\begin{aligned} G_0 &= \{\langle 0,1\rangle, \langle 3,2\rangle, \langle 1,2\rangle, \langle 2,3\rangle\} \\ G_1 &= \{\langle 1,2\rangle, \langle 3,0\rangle, \langle 2,1\rangle, \langle 1,3\rangle\} \\ G_2 &= \{\langle 1,0\rangle, \langle 2,0\rangle, \langle 2,2\rangle, \langle 3,1\rangle\} \\ G_3 &= \{\langle 0,0\rangle, \langle 0,2\rangle, \langle 0,3\rangle, \langle 3,3\rangle\}. \end{aligned}$$

A label matrix L_I facilitates the storing of the group numbers of image pixels. Each location of this matrix contains the value $k, 0 \leq k \leq m - 1$, if the corresponding pixel location in the image I belongs to the group G_k. This label matrix L_I, which is of same size as the image, is the image key. For the example image I_e, this is shown in Fig. 2.

In our proposed watermarking scheme, we manipulate pixel values of groups defined in L_I. This is done following a scheme guided by the bit patterns of the buyer key. The generation of the buyer key is described in Section III, while the watermarking process is given in Section III. From a forger's viewpoint, the main question is the difficulty of guessing the image key. This is important since, with the knowledge of the organization of these groups, it would be easy to decipher the buyer key. On the other hand, if estimation of L_I were difficult, it would be impossible to determine the exact buyer key. This is true even if the forger knows the scheme by which pixel values are manipulated. The following proposition estimates the complexity of that possibility of guessing the image key.

Proposition 1: Consider an image I of size $2^a \times 2^b$. Assume image I is partitioned into $m = 2^n$ groups $G_0, \ldots, G_{m-1}$ each containing equal number of pixel locations 2^{a+b-n}. Then the

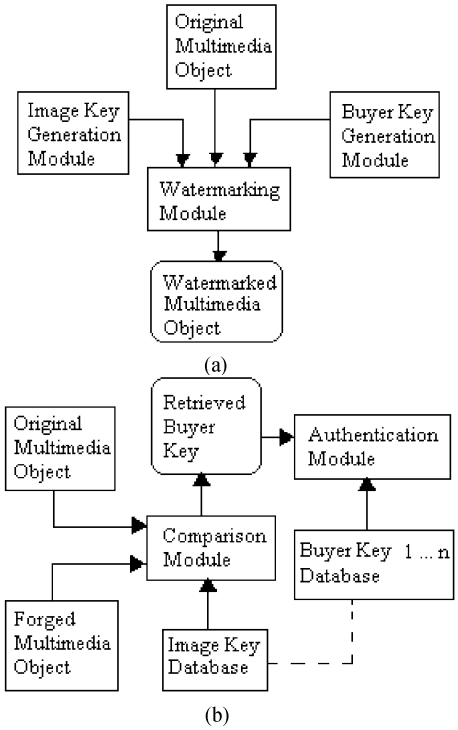


Fig. 3. (a) Watermarking process. (b) Watermark retrieval process.

total number of options to select such groups is greater than $2^{2^{n-1}(a+b-1)}$

Proof: Given the partition of 2^n groups, the number of pixels at each group is 2^{a+b-n}. Thus the total number of choices to select such groups is equal to

$$\begin{aligned}\xi &= \binom{2^{a+b}}{2^{a+b-n}} \times \binom{2^{a+b}-2^{a+b-n}}{2^{a+b-n}} \\ &\quad \times \binom{2^{a+b}-2\times 2^{a+b-n}}{2^{a+b-n}} \\ &\quad \times \cdots \times \binom{2^{a+b}-(2^n-1)\times 2^{a+b-n}}{2^{a+b-n}} \\ &= \prod_{k=0}^{2^n-1} \binom{2^{a+b}-k\times 2^{a+b-n}}{2^{a+b-n}} \\ &> \prod_{k=0}^{2^{n-1}-1} \binom{2^{a+b}-k\times 2^{a+b-n}}{2^{a+b-n}} \\ &> \binom{2^{a+b}-2^{n-1}\times 2^{a+b-n}}{2^{a+b-n}}^{2^{n-1}} \\ &= \binom{2^{a+b-1}}{2^{a+b-n}}^{2^{n-1}} > (2^{a+b-n})^{2^{n-1}} = 2^{2^{n-1}(a+b-1)}.\end{aligned}$$

For an image I, it is now clear that the total number of options in choosing a label matrix L_I is prohibitively large. We select a random label matrix from this set and use it as the image key K_I. Thus the image key has 2^{a+b} locations, each containing an integer value in between 0 to $2^n - 1$. This integer value can be represented using n bits. Thus the total size of the image key K_I is $n2^{a+b}$ bits. Given p image pixels, generation of an image key is an $O(p)$ operation. The image key is stored with the owner of the image, and there is no need to communicate this key while disseminating the multimedia object. Next we discuss the generation of the buyer key.

	0	1	2	3
0	5	7+1	11	233
1	34-1	14+1	79	61
2	123-1	211	47-1	11+1
3	0	254-1	1+1	191

Fig. 4. Example matrix for the watermarked image, I_{ew}.

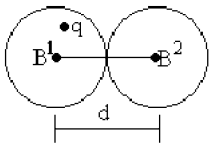


Fig. 5. Spheres representing code words B^1 and B^2 with minimum distance between them is d. The retrieved code word q will be mapped to B^1 following principles of error correction [12].

B. The Buyer Key

Depending on the number of groups $m = 2^n$ in the image key, we take a binary vector of length 2^n. This vector is considered as the buyer key B. Each location of the bit vector B can be accessed as B_k for $0 \leq k \leq 2^n - 1$. Vector B is selected from a set of binary error correcting codes C. The set contains M distinct code words such that Hamming distance between any two code words is at least d. For experimentation, we use the set of 2^n length code words containing $M = 2^{n+1}$ distinct codes with minimum distance $d = 2^{n-1}$ (The Reed Muller Code) [12]. However, we may need error correcting codes with higher minimum distances in some cases. For example, to survive a collusion attack, we need a code with much higher minimum distance than in the other cases. The motivation of selecting buyer keys from a set of error correcting codes will be clear in Section III, where watermark insertion and retrieval issues are discussed.

Assuming that the schemes for generating error-correcting code are known, it would be easy for an attacker to guess the code word set from which the buyer key is selected. Therefore, the actual buyer key used is not directly derived from the error correcting code set that we use. Rather it is a permutation $\pi(B)$ of the code word B where π is kept secret. Note that this permutation $\pi(.)$ is selected randomly, but it is specific for a given image. This permutation provides additional secrecy given that a possible forger may know the error correcting codes but not the image specific permutation. Given a moderate value of the code length m, such possible permutations are m. In subsequent discussions, this transformation is not explicitly mentioned as it has no additional influence on the watermarking and retrieval algorithms described next.

III. WATERMARKING

The overall approach of the proposed scheme is presented in Fig. 3. In the watermarking module [Fig. 3(a)], the original image is spatially divided into a number of blocks based on image key. The image intensity of each block is then modulated depending on the bit values of the buyer key. This process generates the watermarked image.

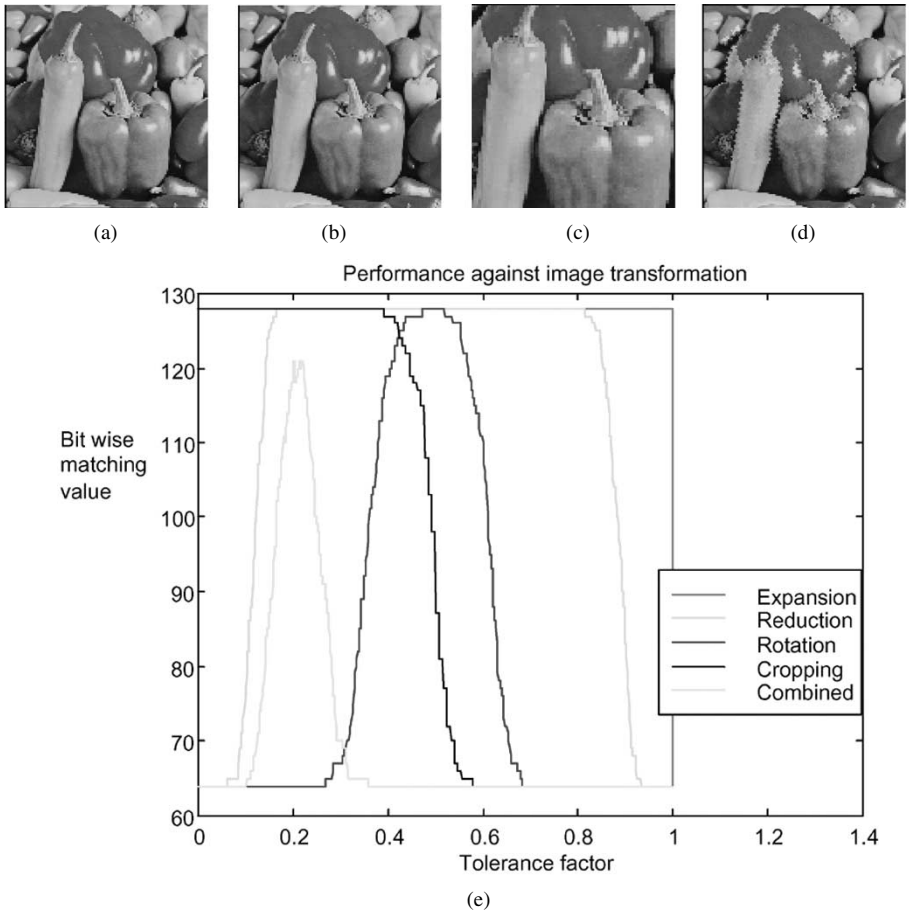


Fig. 6. (a) Original "Peppers" image. (b) Original "Peppers" image after watermarking. (c) Cropped image (enlarged to size 128 × 128). (d) Watermarked image after combined image transformations. (e) Result of buyer key retrieval under expansion, reduction, rotation, cropping and combined image transformations. In all these cases, the retrieval is successful as the bit wise matching values are greater or equal to 75% ≈ 96/128 (in graph, bit wise matching is shown as the number of correct bits out of 128).

TABLE I
IMAGE TRANSFORMATION PARAMETERS, THE WEBER RATIO VALUE THE RANGE OF TOLERANCE FACTOR FOR WHICH THE BIT WISE MATCHING VALUE ≥97

Transformation Type	Transformation parameter	Weber ratio (after attack)	Tolerance factor range for bit wise matching value ≥ 97
Expansion	1.27×	0.008	0~1
Reduction	0.79×	0.087	0~0.87
Rotation	13°	0.173	0.35~0.63
Cropping	~73% of original area	0.191	0~0.5
Combined	0.83×, 13° rotation, ~80% of original area	0.442	0.17~0.28

In the retrieval phase [Fig. 3(b)], the original and the watermarked (may be forged) images are compared block by block. The block information is obtained from the image key. Depending on the extent of intensity modification in each block a probable buyer key is generated. This key is then mapped to exact buyer key by correcting the errors using the theory of error correcting codes [12]. In Section III-A, we present the watermark insertion and the buyer key retrieval algorithm.

A. Insertion and Retrieval of Watermark

The process of generating watermarked image I_w from the original image I is described in this section.

Algorithm 1

1) For $0 \le i \le 2^a - 1,\ 0 \le j \le 2^b - 1$
   a) Let $I(i,j)$ be the pixel value of the image I at the pixel location $\langle i,j \rangle$.
   b) Assume that $\langle i,j \rangle$ belongs to the group G_k, i.e., $L_I(i,j) = k$.
   c) If $B_k = 0$, then $I_w(i,j) = I(i,j)$.
   d) Else if $B_k = 1$, then $I_w(i,j) = I(i,j) + \beta(i,j)$

For 1(d), we denote the sum $\sum_{\langle i,j \rangle \in G_k} (I_w(i,j) - I(i,j)) = \sum_{\langle i,j \rangle \in G_k} \beta(i,j) = \delta_k$. We show that the δ_k values play an important role in this digital watermarking technique. We consider

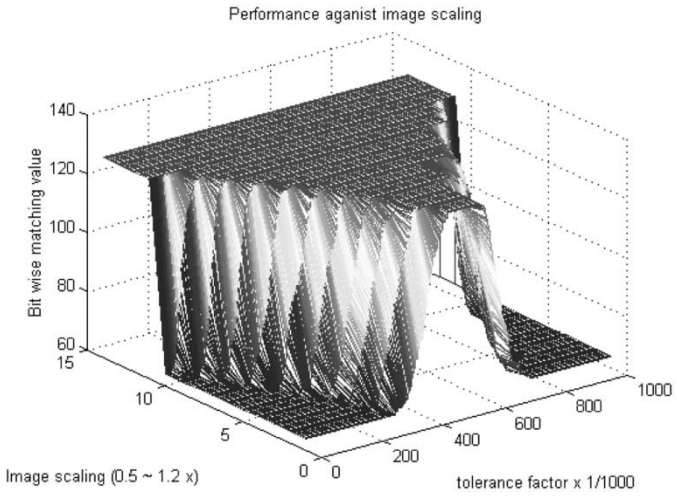


Fig. 7. Performance against image scaling. The buyer key recovery is demonstrated for scaling range 0.5 to 1.2 times the original size.

that the $\beta(i,j)$'s are either all positive or all negative corresponding to a group G_k. Whether $\beta(i,j)$ will be taken positive or negative is based on a uniformly distributed random variable. Thus, the values of δ_k may be either positive or negative. Also it is important to decide the values of $\beta(i,j)$ such that the quality of the image is not degraded. We follow the principles of the Weber ratio (WR) in selecting the value of $\beta(i,j)$ [6]. To maintain perceptual quality, $WR = (\sum I_w - \sim I) / \sum I$ is taken as less than or equal to 2% of the original image intensity value. Also, note that original image intensities are not at all changed in case $B_k = 0$ (step (c) of Algorithm 1). The image intensities (for the 8-bit case) are restricted between 0 and 255 (in which decreasing a 0-valued pixel and increasing a 255-valued pixel are prohibited).

Let us discuss the situation in terms of the example image I_e used in Section II. From I_e, we construct the watermarked image I_{ew} for $0 \le i \le 3, 0 \le j \le 3$ and the buyer key $B = 1010$. For $B_k = 1$, in pixel groups G_0 and G_2, we take $\beta(i,j) = 1$ and $\beta(i,j) = -1$, respectively. This is shown in Fig. 4. In this case, $\delta_0 = 4, \delta_2 = -4$, and $\delta_1 = \delta_3 = 0$. The objective of watermark retrieval is to get back the buyer key B from the watermarked image I_{ew} given the original image I_e and the image key K_I. The following algorithm is used for this purpose.

Algorithm 2

1) For $0 \le k \le 2^n - 1$
   a) Initialize values $\sigma_k = 0$.
   b) Initialize bit values $q_k = 0$.
2) For $0 \le i \le 2^a - 1, 0 \le j \le 2^b - 1$
   If $\langle i,j \rangle$ belongs to the group G_k, then $\sigma_k = \sigma_k + I_w(i,j) - I(i,j)$.
3) For $0 \le k \le 2^n - 1$
   a) If σ_k and δ_k are equal with the value 0, then $q_k = 0$.
   b) If σ_k and δ_k are equal with nonzero value, then $q_k = 1$.
   c) Report q as the buyer key B.

For p pixels, the retrieval of the buyer key is executed in $O(p)$ time. If User 1 buys a watermarked image I_w from the owner and then resells I_w to User 2, then from I_w (no matter whether the owner gets it from User 1 or User 2), the owner can easily compute the buyer key and identify User 1. However, the actual scenario is more complicated. User 1 can utilize intentional processing within the image such that the watermarking scheme is disturbed. In such a situation, the owner cannot easily identify the buyer key as given in the above algorithm. This situation is discussed next.

B. Identifying Buyer Key From Attacked Watermarked Image

The following algorithm allows one to compute the buyer key successfully from the attacked watermarked image $I_{w\#}$. Let us first describe the algorithm.

Algorithm 3

1) For $0 \le k \le 2^n - 1$
   a) Initialize values $\sigma_k = 0$.
   b) Initialize bit values $q_k = 0$.
2) For $0 \le i \le 2^a - 1, 0 \le j \le 2^b - 1$
   If $|I_{w\#}(i,j) - I(i,j)| > |\beta(i,j)|$, then $I_{w\#}(i,j) = I(i,j) + \beta(i,j)$.
3) For $0 \le i \le 2^a - 1, 0 \le j \le 2^b - 1$
   If $\langle i,j \rangle$ belongs to the group G_k, then $\sigma_k = \sigma_k + I_{w\#}(i,j) - I(i,j)$.
4) For $0 \le k \le 2^n - 1$
   a) If $|\sigma_k| \le c_k|\delta_k|$, then $q_k = 0$.
   b) Else if $|\sigma_k| > c_k|\delta_k|$, then $q_k = 1$.
5) Compute the code word q' closest to q.
6) Report q' as B.

Step 2 in Algorithm 3 is used for pruning the watermarked (and possibly forged) image intensities against impossible values. Given the watermarking scheme and the $\beta(i,j)$ values, the difference between $I(i,j)$ and $I_w(i,j)$ is known. Thus, if we receive an attacked watermarked image $I_{w\#}$ such that $|I_{w\#}(i,j) - I(i,j)|$ is greater than $|\beta(i,j)|$, then it is clear that this cannot be a true value for the watermarked object. In this situation, the value of $I_{w\#}(i,j)$ should be replaced by $I_{w\#}(i,j) = I_w(i,j) = I(i,j) + \beta(i,j)$ before further processing.

Now it may so happen that during image transformation or forging attempts, $I_w(i,j)$ is changed to $I_{w\#}(i,j)$ in such a way that it may be difficult to interpret the values of $\beta(i,j)$ for $I_w(i,j)$. Let us explain this with the watermarked image example I_{ew} of Fig. 4. After watermarking for the buyer key $B = 1010$, $\delta_0 = 4$ for the pixel group G_0 ($\beta(i,j)$ is 1 for G_0 as 0th bit of the buyer key is 1). If due to image transformation or forging attempts, the individual pixel values of group G_0 were distorted in such a way that δ_0 becomes 0, we lose the information that originally $\beta(i,j)$ was 1 for this group. Consequently, after attack the corresponding bit for the buyer key would be incorrectly identified as 0 instead of 1. So for a given group G_k, we get $0 \le |\sigma_k| = \sum_{\langle i,j \rangle \in G_k} |I_{w\#}(i,j) - I(i,j)| \le |\delta_k|$. The decision at this point is between interpretation of σ_k as 0 or as δ_k. If we interpret σ_k as 0, then the corresponding k^{th} bit posi-

TABLE II
BITWISE MATCHING VALUE AGAINST A RANGE OF TOLERANCE FACTOR AFTER STIRMARK 3.0 [14] ATTACK. SUCCESSFUL RECOVERY IS POSSIBLE FOR TOLERANCE FACTOR RANGE 0.5–0.7

$c_k \rightarrow$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Bitwise match	82	90	93	94	98	100	97	94	94	88

TABLE III
BITWISE MATCHING VALUE AGAINST A RANGE OF TOLERANCE FACTOR AFTER JPEG COMPRESSION ATTACK (GF IS THE QUALITY FACTOR OF THE JPEG COMPRESSION). BUYER KEY IDENTIFICATION IS POSSIBLE FOR ALL UNDERLINED BITWISE MATCHING VALUE (≥ 97)

$c_k \rightarrow$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Bitwise match (qf=10%)	74	77	81	84	95	95	98	98	88	83
Bitwise match (qf=20%)	92	103	114	120	117	110	109	100	94	88
Bitwise match (qf=30%)	91	112	123	125	126	123	115	108	94	78
Bitwise match (qf=50%)	107	124	126	128	128	128	123	112	97	80
Bitwise match (qf=70%)	121	126	128	128	128	128	127	118	100	76
Bitwise match (qf=90%)	126	128	128	128	128	128	127	125	115	82

tion in the buyer key q is 0; else, we interpret the bit as 1. The value of the tolerance factor c_k as in step 4 of Algorithm 3 plays an important role in this respect. In Section IV, we expand on this point.

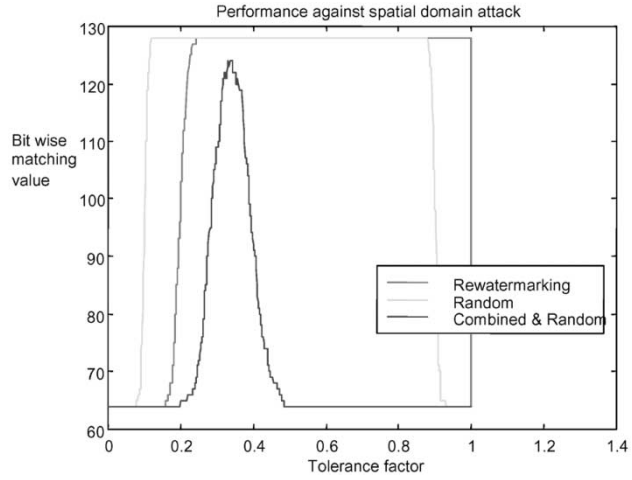
C. Analysis of Watermark Retrieval Process

Exact determination of c_k is not required. Instead, we use a range of c_k values. The intensity profile of each image block may change differently due to an attack; however, in no case should the value cross the perceptual limit as that may distort the quality of the multimedia object. The range of c_k quantifies the extent to which the block intensity profile is changed with respect to the original. Such change is guided by the Weber ratio (2% of the original value). We present our retrieval result as the number of bit wise matches between B and q against this range of c_k. Let us analyze the properties responsible for correct retrieval of watermark.

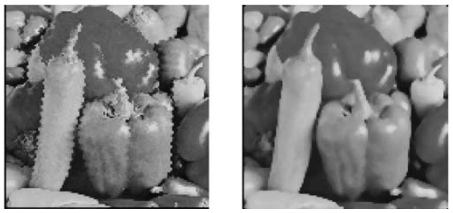
The method is applied over some disjoint subsets of the image specified by the image key. To select the image key, as per Proposition 1, the attacker has a choice out of *minimum* $2^{2^{n-1}(a+b-1)}$ possibilities when an image of size $2^a \times 2^b$ is divided into $m = 2^n$ groups. So, it is clear that guessing the image key is impossible. Also the exact buyer key is not known without the permutation $\pi(B)$ for a particular buyer code B as mentioned in Section II-B.

Now, let us analyze the proposed watermarking process. The watermark increases or decreases the intensity values of an image group defined by the image key. Since the attacker is aware of this process, he/she might try to invoke the reverse process to eliminate the watermark or rewatermark the watermarked image with the intention of destroying the original watermark. In both cases, the attacker will be regrouping the pixels followed by an increment or decrement of pixel intensities. Without the knowledge of the exact image key, for a subgroup G_k, to some extent the increment or decrement due to attack will be nullified with respect to decrement or increment of original pixel intensities for watermarking.

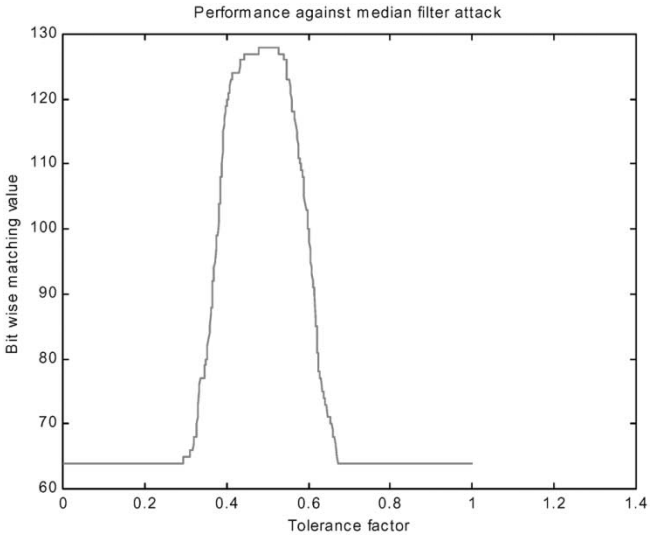
It may very well happen that the attacked image $I_{w\#}$ is such that there are some errors in deciding the bits of q. Such possibility is already discussed in Section III-B with an example. We already know that the buyer key B is chosen from a set of error correcting codes. If q is already a code word, then we choose $q' = q$. Else we try to find a code word q' closest to q and report q' as the buyer key B. Since the minimum distance between the code words is d, even if there are $\lceil d/2 \rceil - 1$ errors in determining q, i.e., the Hamming distance between q and B is at most $\lceil d/2 \rceil - 1$ at the time of finding the closest code word q', these errors can be corrected [12]. Thus we obtain the proper buyer key $B = q'$. On the other hand, if the number of errors exceeds $\lceil d/2 \rceil - 1$, then an incorrect buyer key will be estimated and the scheme will fail. This process can be visualized in Fig. 5. Let each code word represents the center of a sphere having radius $\lceil d/2 \rceil$. The center-to-center distance between two such spheres is d. The retrieved bit pattern q is mapped to the code word representing the center of the sphere within which q is lying. For a given code word, the scheme fails if q is outside the sphere. This could only be possible if at least in $\lceil d/2 \rceil - 1$ regions, estimation of σ with respect to known δ is wrong as specified in step 4 of Algorithm 3. This is possible only if attacker can guess



(a)



(b) (c)

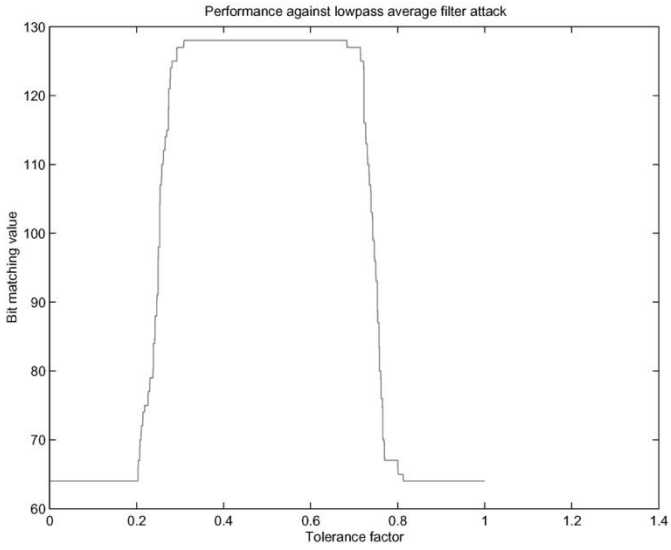


(d)

Fig. 8. (a) Results for spatial attacks of rewatermarking, random intensity manipulation and combination of scaling, rotation, cropping and random attack (in graph, bit wise matching is shown as the number of correct bits of 128). (b) Resultant image after combined and random attack on spatial domain watermarked image. (c) Result of median filtering on the watermarked "Peppers" image. (d) Result of buyer key retrieval in the case of a median filtered watermarked image (in graph, bit wise matching is shown as number correct of 128).

at least $\lceil d/2 \rceil$ image groups defined in image key. However, it is not possible (highly unlikely) to guess $\lceil d/2 \rceil$ groups as the number of options is exponentially high (follows from the proof of Proposition 1).

For high value items, it is natural that we need to provide additional security. Also, it is expected that fewer copies will be sold for an expensive item. This means we need comparatively fewer



(e)

Fig. 8. (*Continued*.) (e) Result of buyer key retrieval in the case of lowpass average filter attack watermarked image (in graph, bit wise matching is shown as number correct of 128).

buyer keys, which helps in selecting a larger value for d. For example, in the case of only two buyers for a multimedia object with 8-bit long buyer key, the Hamming distance between two buyer keys could be much larger compared to a situation where there is, for example, 64 buyers. According to the principles of error correction, even if there is bitwise mismatch (between the retrieved bit pattern q' and a buyer key in the database) in at most $\lceil d/2 \rceil - 1$ positions, the error could be corrected. Since, the distance d is large in case of high value item, there will be better error correction for high value items. That is, even if the attack is severe in distorting the intensity profile of several image blocks of a high value item, the correction capability is also increased.

Nonlinear geometric and compression attacks can change the block intensity profile significantly. Therefore, additional features are necessary to prevent such attacks. In Section IV, we address this issue.

D. Resistance to Nonlinear Attacks

In this section we show that a slight variation in the image key organization and corresponding changes in watermark recovery process can prevent nonlinear attacks even if the attacker is aware of our watermark insertion strategy. Note however, that the buyer key definition needs no change.

Given an image I of size $2^a \times 2^b$, we can first consider that the image is subdivided into $m = 2^n$ groups denoted by $G_0, G_1, \ldots, G_{m-2}, G_{m-1}$. Let us further define a set of smaller *contiguous* pixel units of size $2^\alpha \times 2^\beta$ within each of these groups. Therefore, there are $\tau = 2^{a+b-n-\alpha-\beta}$ units within each group. For a particular group G_k, these units are denoted as $G_{k,0}, G_{k,1}, \ldots, G_{k,\tau-2}, G_{k,\tau-1}$. During watermarking, pixel intensities of contiguous units of G_k are either all modified or all constant depending on bit values of buyer key similar to Algorithm 1. During recovery process, we first check the status of units that were not modified at all during watermarking. Let us call them 0-bit units. The spatial position of the 0-bit units may be changed due to nonlinear geometric

TABLE IV
PARAMETERS IN CASE OF SPATIAL DOMAIN ATTACK, THE WEBER RATIO VALUE AND THE RANGE OF TOLERANCE FACTOR FOR WHICH BIT WISE MATCHING VALUE ≥ 97 ARE PRESENTED

Transformation type	Transformation parameter	Weber ratio (after attack)	Tolerance factor range for bit wise matching value ≥ 97
Rewatermarking		0.002	0.2~1
Random		0.079	0.1~0.9
Combined and Random	0.83×, 13° rotation, ~80% of original area	0.541	0.3~0.4
Median filtering	3x3 mask	0.012	0.38~0.65
Average filtering	3x3 mask	0.07	0.2~0.8

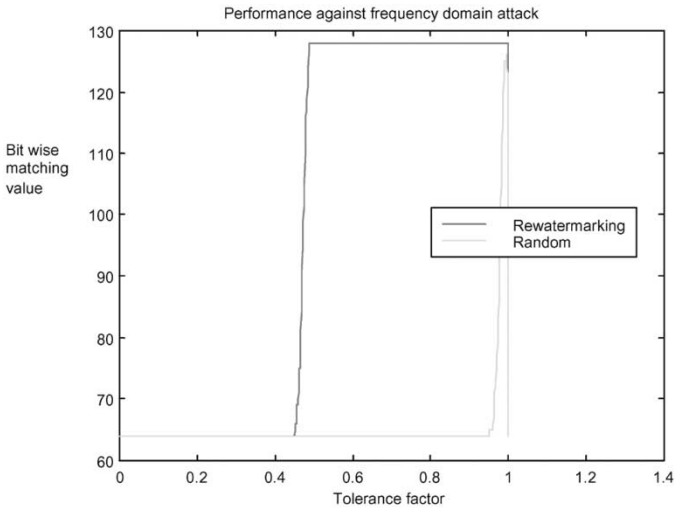
attack. The new positions of the 0-bit units are determined using a correlation scheme within a search window in the attacked image. For 0-bit units, the average change in intensity due to attack is estimated. Intensity of each unit within a group is normalized with respect to the average change in intensity due to attack. Then the buyer key is calculated following the steps similar to Algorithm 3. The modified watermark recovery algorithm is presented next.

1) For every 0-bit unit of group G_k in I, find the best matched unit in $I_{w\#}$. The match is restricted within a window in $I_{w\#}$. The center of the window is spatially identical to the top-left corner of the 0-bit unit of G_k in I. The match measure is defined as the sum of absolute differences between every corresponding pixel within the unit. Let $\varphi_{x,y}$ is the total change in intensity for a particular 0-bit unit. Note that the intensity change $\varphi_{x,y}$ is purely due to the attack as 0-bit units are not modified while watermarking.
2) For a total of p 0-bit units in the group G_k, the average amount of change in pixel intensity in G_k due to attack is $\gamma = \sum_p \varphi_{x,y}/p2^{\alpha+\beta}$. Each pixel of the group G_k of the attacked image is then pruned by the following amount: $I'_{w\#}(i,j) = I_{w\#}(i,j) - \gamma$.
3) The buyer key is then identified following steps 4 to 6 as in Algorithm 3.

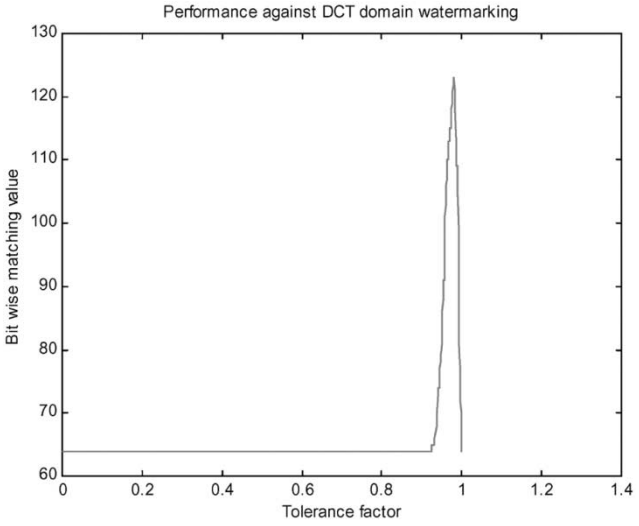
We have used small contiguous units of size 4×4 or 8×8 for every group. Most of the small units preserve intensity information even after geometric attack even though spatially relocated to a new position. Their new locations are identified using the window matching algorithm as stated above. The window based match measure finds the correspondence even after spatial transformations. The small contiguous units keep the intensity information in case of block based JPEG transformations. In Section V, we simulate the insertion and retrieval of watermark in multimedia objects and show the strength of the methodology under a variety of image transformation and forging attacks including the Stirmark and JPEG compression attacks.

IV. SIMULATION

We have used Reed Muller codes [12] to generate 2^n length buyer keys with the minimum distance between any two codes being 2^{n-1}. There are 2^{n+1} such codes. For experimentation, we have taken $n = 7$. This makes the length of the buyer key 128 bits and can provide maximum number of 256 (2^{n+1})



(a)



(b)

Fig. 9. (a) Results for the frequency domain attack on an image where watermarking is performed in the spatial domain (in graph, bit wise matching is shown as number correct of 128). (b): A rewatermarking is performed within the DCT domain of the watermarked image following [2] (in graph, bit wise matching is shown as number correct of 128).

TABLE V
FREQUENCY-DOMAIN ATTACK TYPE, WEBER RATIO VALUE AND THE RANGE OF TOLERANCE FACTOR FOR WHICH THE BIT WISE MATCHING VALUE ≥ 97 IS SHOWN

Transformation type	Weber ratio	Tolerance factor range for bit wise matching value ≥ 97
Rewatermarking	0.039	0.5~1
Random	0.102	0.97~1

distinct code words. Note that as mentioned in Section II-B, this code word set is subjected to a random permutation $\pi(\cdot)$ specific to an image. Bit wise matching value of the 128 bits between the buyer key and the retrieved bit pattern from the attacked image reveals the identity of the buyer. The scheme can correct a maximum of 31 $(2^{n-2} - 1)$ bit errors. So, bit wise matching between retrieved pattern and the buyer key in the database in at least 97 $(= 128 - 31)$ positions should ensure complete decoding of the buyer key. Throughout the paper, we

TABLE VI
COMPARISON OF REPETITIVE REWATERMARKING ATTACK. THE PARAMETER "C" REPRESENTS THE CORRELATION BETWEEN THE INSERTED SIGNATURE AND THE EXTRACTED SIGNATURE FOR [2]. "B" IS THE MAXIMUM BIT WISE MATCHING VALUE (IN BITS) BETWEEN BUYER KEY AND THE RECOVERED BIT PATTERN WHILE "W" STANDS FOR WEBER RATIO

Watermarking	#1 Watermark		#2 Watermark		#3 Watermark		#4 Watermark		#5 Watermark	
Using [2]	C	W	C	W	C	W	C	W	C	W
	0.61	0.014	0.42	0.031	0.32	0.057	0.25	0.082	0.19	1.21
Proposed approach	B	W	B	W	B	W	B	W	B	W
	128	0.002	128	0.009	123	0.012	110	0.018	101	0.025

have described success of the scheme whenever there is at least 97 bits that match or 75% matching. In general, for an n -bit buyer key, bit wise matching in a minimum $(2^n - 2^{n-2} - 1)$ positions ensures correct recovery of the buyer key.

In the experiments, the watermark is added in the spatial domain, and its performance is tested against a set of possible image transformations and simulated attack. A major result of the paper is given in demonstrating the strength of the proposed approach against Stirmark and JPEG compression attacks. Referring to the second generation attacks proposed by Voloshynovskiy *et al.* [16], we have shown that most of the attacks mentioned in their paper can be repelled using our method. The proposed approach survives all of the removal attacks including image filtering attacks using both median and average filters. For a collusion attack, the variation to the proposed approach including the limitation and open problems is discussed in Section V. The demonstration of the method against geometrical attacks is shown using rotation, scaling and warping transformations simulated through Stirmark 3.0. Extensive testing with different forms of image cropping is described. Through proposition I, we have already explained that cryptographic attacks such as a brute force key search oracle is simply impossible in this case, despite the fact that we are assuming that attacker knows our watermarking scheme and consider that also to be a valid attack. This brings our approach close to the last category of attacks mention in [16], namely the protocol attack. A key contribution of this method is that even if the attacker tries to rewatermark the image, the attacker cannot delete the buyer specific key.

Within a 128×128 image, the image is divided into 128 different image blocks. The results are presented in the form of graph in which bit wise matching values are plotted against the tolerance factor described in Section III. In the graph, bit wise matching value greater or equal to 97 matches $(>75\%)$ within the given range of tolerance factor ensures complete recovery of the buyer key. Throughout the experiment, we have set the $\beta(i, j)$ values equal to $\beta = 1$, as it is the minimal level of intensity modification of the pixels in spatial domain. Naturally, this equates to the most favorable situation for the attacker and the most challenging scenario for retrieval. In the subsequent discussions, we show that our method survives satisfactorily in retrieving the buyer key. Knowing the value of β, we know the upper and lower limits of each pixel after watermarking. So, as described in Step 2 of Algorithm 3, the attacked and/or transformed image is filtered yielding unexpected pixel values.

TABLE VII
PERFORMANCE OF SPATIAL WATERMARKING SCHEME AGAINST DCT DOMAIN ATTACK [2]

Transformation Type	Transformation parameters	Weber ratio (after attack)	Tolerance factor range for bit wise matching value ≥ 97
DCT domain attack	Following [2]	0.0593	0.94~1

To evaluate the performance of our scheme, a wide range of simulations have been enacted, and the retrieval process proves successful in all cases. We present here a representative set of results that highlight the contribution of the method. We have implemented the algorithms in Matlab (Mathworks, Natick MA) on an Intel PIII 800 MHz CPU. For 128-bit buyer keys applied on 128×128 image, the watermarking process is almost instantaneous while the buyer key retrieval takes approximately 0.3–0.7 s, depending on the tolerance factor used.

A. Cryptographic Robustness in Spatial Domain

The different image transformations tested are scaling, rotation, cropping and a combination of all the aforementioned transformations. The original "Peppers" image of size 128×128 is shown in Fig. 6(a). The watermarked "Peppers" image is shown in Fig. 6(b). The buyer key retrieval result is shown in Fig. 6(e). For this simulation the watermarked image is separately subjected to 127% expansion, 79% reduction and a rotation of 13°. The cropped image for testing is given in Fig. 6(c). In case of combined transformation, the sequences of transformations include 83% reduction (scaling) followed by 17° rotations and cropping that maintains approximately 80% of the original watermarked image. The resultant transformed image is shown in Fig. 6(d). Table I details the image transformation parameters, the Weber ratio value and the range of tolerance factor c_k for which the bit wise matching value is greater or equal to 75%. Note that for the image in Fig. 6(d), we could successfully recover the buyer key even if the quality degrades beyond the perceptually acceptable limit.

To investigate the effect of the scaling transformation on watermarking, a watermarked image is scaled in size between 0.5 to 1.2 times in steps of 0.05. The corresponding bit wise matching values are shown in Fig. 7. The range of c_k is 0.001 to 1 in steps of 0.001. Note that as the scaling factor approaches 1 (and higher), there are fewer pixels interpolated or introduced due to scaling, and the error in buyer key positions is decreasing for a wide range of tolerance factors.

Analysis of cropping attacks is very important and linked to the strength of the process against collusion attack. We have simulated two different kinds of cropping attacks: regular and random types. In the regular cropping attack, the image is cropped from any one corner of the image (bottom right corner for our experiments) and the extent of cropping is gradually increased, until the approach fails to detect the owner. Note that in all these cases, the cropped region is first replaced by the corresponding portion of the original image. In case of random cropping attack, portions of the watermarked image are randomly replaced by the corresponding portions of the original image. Overall, the buyer key is successfully recovered up to 37% cropping in case of regular cropping (that is 63% of the watermarked image remains after attack). The strength increases in the case of random cropping where authentication is successful up to 59% of random replacement of the watermarked image pixel with the original one. The reason for better cropping resistance in the random case is due to the fact that the image key organization is spatially random in the image space.

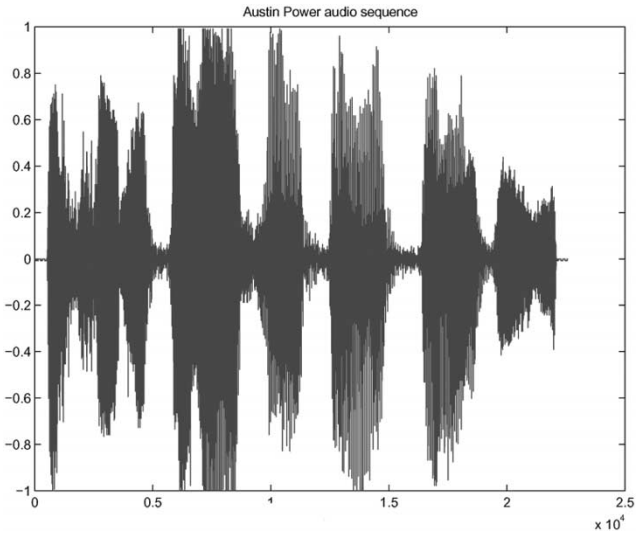
For further analysis, we have tested the variation of the proposed watermarking scheme of Section III-D both for nonlinear geometric attacks through Stirmark 3.0 [14] and compression attacks using JPEG. In both cases, successful recovery of buyer key is possible as demonstrated in Tables II and III, respectively. We have used four contiguous pixel units of size 4×4 for every group defined by the image key. The search for a matching unit in the attacked image is restricted within 16×16 window with its center overlapping with top-left corner of the unit. The buyer key length is also 128 in this case. Because of smaller unit size, bit wise matching value in excess of 97 $(>75\%)$ is obtained for a wide range of c_k, from 0.5 to 0.7. For the JPEG compression attack, the simulation is performed for a wide variety of quality factors ranging from 10% to 90%. As expected, the performance of buyer key recovery through extensive bit wise matching between recovered key and the buyer key is increased as the quality factor improves to 90%.

B. Performance Against Spatial Domain Attack

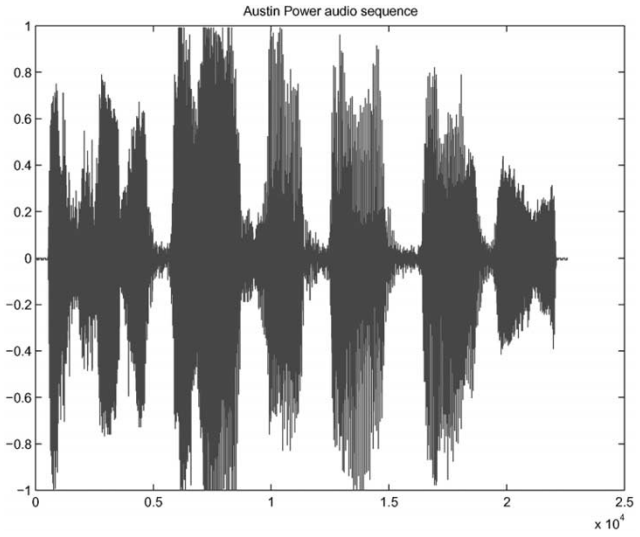
Attempts to destroy the watermark in the spatial domain are the most common type of attack in digital watermarking. We have simulated four such conditions, and the performance of our proposed scheme against such attacks is demonstrated.

a) In the first case, rewatermarking is performed on the watermarked image. The parameters for the process are exactly identical as the original watermarking except that different image and buyer keys are used for re-watermarking. This is in line with our assumption that the attacker is aware of the watermarking algorithm. The result of watermark retrieval is shown in Fig. 8(a).

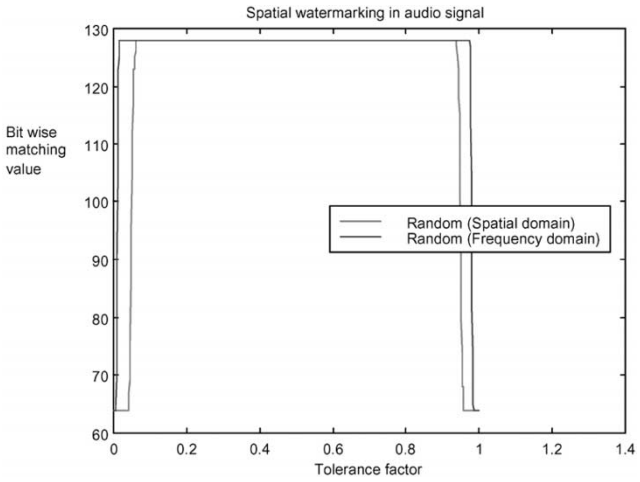
b) The next test is to corrupt the intensity values of the watermarked image by either increasing or decreasing intensity by a small amount. Increment or decrement operations are enacted randomly in this case, and the image distortion is within the permissible limit given by the Weber ratio. The result of watermark retrieval for this scenario is shown in Fig. 8(a).



(a)



(b)



(c)

Fig. 10. (a) Original "Austin Powers" audio sequence. (b) Watermarked "Austin Powers" audio sequence on the original signal as shown in (a). (c) Results for random attack both in spatial and frequency domains on the watermarked audio signal (in graph, bit wise matching is shown as number correct of 128).

c) The random attack, similar to situations in (b), is implemented on the watermarked image already subjected to combined image transformations of scaling, rotation and cropping. These parameters are same as in the case of Fig. 6(d). The result of watermark retrieval for this combined attack is shown in Fig. 8(a), while the resultant image is shown in Fig. 8(b).
d) Spatial domain filtering is also a potential attack on a watermark. Here, median filtering and average filtering are considered. The result of watermark retrieval after median and average filtering attacks are shown in Fig. 8(d) and (e), respectively.

In all these cases, the proposed watermarking scheme is successful as bit wise matching values greater than 75% are achieved. Exact bit wise matching (100%) is achieved in the cases of rewatermarking and random attacks. For a combination attack as described in (c), the maximum bit wise matching value is 123 of 128 (96%), sufficient for recovery of buyer key. Moreover, buyer identification is possible for this case even though the image quality is perceptually not acceptable after attack. Numerical results showing image transformation parameters, Weber ratio and the range of tolerance factor for which the bit wise matching value is $\geq 75\%$ are presented in Table IV.

After a median filtering attack on the watermarked "Peppers" image using a 3×3 window, the resultant image is shown in Fig. 8(c). The performance of watermark retrieval is shown in Fig. 8(d). The retrieval is successful in this case as bit wise matching $\geq 75\%$ is obtained for a wide range of tolerance factors (0.38–0.65), including a peak matching value of 100% showing the exact match. For a low pass filtering attack using a 3×3 average filter, the successful buyer key retrieval results are shown in Fig. 8(e), along with the corresponding bitwise matching values in Table IV.

As mentioned in the introduction, the attack can be extended in frequency domain as well. In Section IV-C, we simulate a set of attacks involving frequency components of the digitized image.

C. Performance Against Frequency Domain Attack

The simulation of forging attempt is further extended to frequency domain. The watermarked image is transformed to frequency domain using the Fast Fourier Transform (FFT). The frequency domain image is subjected to two different attacks:

a) Rewatermarking: A separate watermark is inserted in the frequency domain following watermarking principles explained in Algorithm 1.
b) Random: Similar to the random attack in the spatial domain, coefficients in frequency domain are randomly manipulated. The amplitude is changed to a maximum $\pm 10\%$ of the original value.

After these attacks, the inverse FFT is applied and the image is subjected to the watermark retrieval process. The bit wise matching values are shown in Fig. 9(a). In both the rewatermarking and random cases, successful retrieval of buyer key is possible as we have achieved bit wise matching values greater than 75%. In the case of the random attacks, however, the tolerance factor range is small compared to others for which bit wise matching value is $\geq 75\%$. This shows the strength of the process, as this particular attack is quite destructive. The corresponding watermarking parameters are shown in Table V.

D. Comparison With Respect to Spread Spectrum Based Watermarking

Other than testing with Stirmark benchmark [14], we have compared the performance of our approach against the widely cited spread spectrum based watermarking described in [2]. The comparison is done in two different aspects. First, we have studied the rewatermarking attack extensively. As mentioned earlier, we assume that the attacker knows the watermarking process. So we would like to compare the performance of watermark recovery using our approach and the approach in [2] after repetitive rewatermarking. We have repeated the (re)watermarking process five times as a possible attack on the already watermarked image. For spread spectrum based attack, the rewatermarking is done following [2] using five different signatures, and the rewatermarking is halted after five iterations as the quality of the image degrades beyond the perceptual limit in this case. We have used the implementation of watermarking and recovery algorithm of [2] publicly available from [13]. The comparison is shown in Table VI. We can see as the number of rewatermarking attempts increases, the correlation between inserted signature and the extracted signature goes as low as 0.19. However, using our approach, the buyer authentication is done using the exact match between the buyer key and the retrieved bit pattern. In all cases, bit wise matching value between the retrieved bit pattern and the buyer key is well above the 75% required for successful buyer authentication. From the Weber ratio measure, it is also seen that the proposed methods holds advantages in preserving image quality, as compared to [2].

The comparison is further extended using Stirmark benchmark 3.0. We have used default parameters of Stirmark 3.0. For watermarking using [2], the average correlation value between the inserted signature and the extracted signature after Stirmark 3.0 default attack is 0.24. This correlation decreases beyond 0.2 in case affine distortion parameters are enhanced from the default values. Using the window based proposed variation in Fig. 6(a), the corresponding bit wise similarity value ranges between 81–86%, which is sufficient for buyer authentication.

To further assess the robustness of this method, we have considered watermarking scheme proposed in [2] as a valid attack on our watermark. In this case, DCT components are modified to a maximum of 5% of their original value. Then, the watermark is recovered from inverse DCT image. The recovered bit pattern has bit wise matching value $\geq 75\%$ where the peak matching value is 96%. This ensures complete identification of buyer. The result is shown in Fig. 9(b). The numerical details are given in Table VII.

This concept of spatial domain watermarking is extended for audio signals as well. The performance of the watermark retrieval in the audio case is detailed next.

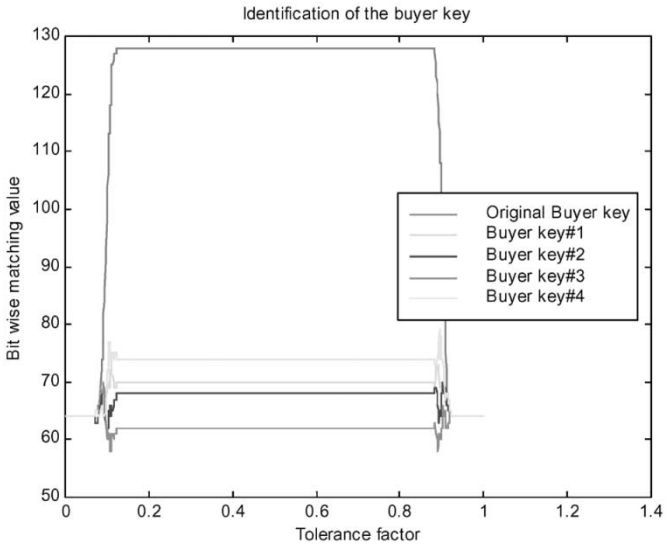


Fig. 11. Exact identification of buyer key from a set of five buyer keys is shown (in graph, bit wise matching is shown as number correct of 128).

E. Attack on Watermarked Audio Sequence

We have used the proposed watermarking technique on a variety of multimedia objects including audio and video sequences. Watermarking of video images is identical to the digital image watermarking described in Section III. For audio, the extension is straightforward, and an example is provided. A representative watermarked audio signal from the film "Austin Powers" is shown in Fig. 10(b). The original sequence is shown in Fig. 10(a). This is a sequence of approximately 2.4 seconds duration (file size 44.1KB). The signal is subjected to random amplitude attack both in spatial and frequency domain. In both cases successful recovery of watermark could be achieved. The amplitude mean square error after watermarking is kept under a maximum of 0.05% of the original amplitude value while after attack it went up to 0.09% in the case of frequency domain attack. The recovery result is shown in Fig. 10(c). The buyer key pattern is retrieved exactly for a wide range of tolerance factor (0.03–0.96).

In the next section, we show the accuracy of our technique in identifying the proper buyer from a group of buyers.

F. Authentication of Buyer Key

We once again refer to the step 5 of Algorithm 3. In this step, we find the correct code word q' closest to q. Our hypothesis is that the selection of an incorrect buyer key is improbable withing the complete range of tolerance factors. We substantiate this with the following experiment. After watermarking, the image intensities are marginally increased or decreased in random spatial locations similar to random attack explained in Section IV-B. Apart from the original one, we select four more code words from the same error correcting code set. Fig. 11 shows that while varying the tolerance factor c_k, the original code word provides the highest bit wise matching with the retrieved bit pattern. For the rest of the code words, bit wise matching values never reach the 75% matching necessary for buyer key authentication.

TABLE VIII
AUTHENTICATION OF BUYER KEY IN CASE OF COLLUSION ATTACK

Retrieved pattern when checked with ↓	Buyer key #1	Buyer key #2	Buyer key #3	Buyer key #4
Image key #1	123	64	64	64
Image key #2	63	124	67	63
Image key #3	64	64	123	64
Image key #4	65	64	64	122
Image key #5	64	64	64	64

This experiment reveals the possibility of collusion attacks. The algorithm that prevents some of these attacks and the performance of our approach in that context are described in Section V.

V. COLLUSION ATTACK

Robustness to collusion attack is an open and challenging problem. It is already shown that for a reasonable model of digital watermarks, if the document length is n, then $O\left(\sqrt{n/\ln n}\right)$ adversaries can defeat any watermarking scheme [5]. The scheme presented so far is vulnerable to collusion attack if we use same value of $\beta(i,j)$ for all the buyer keys. Also we have assumed that the attackers are aware of our watermarking process. By comparing pixel values of the same location in more than one watermarked image, it is possible to guess the true value of the image pixels. So we propose two specific modifications to the watermarking and retrieval algorithms presented in Algorithm 1 and 3 respectively.

Modification 1: Instead of a fixed $\beta(i,j)$, corresponding to the image I we define a range of $\beta(i,j)$ values: $\mu^-(i,j) \le \beta(i,j) \le \mu^+(i,j)$ $\mu^-(i,j) < 0, \mu^+(i,j) > 0$ such that $I(i,j)+\beta(i,j)$ does not make any perceptually significant change for the complete range. That is, we constitute a matrix Γ of size identical to the image size. Each element of Γ is denoted by $\Gamma(i,j)$ where $\mu^-(i,j) \le \Gamma(i,j) \le \mu^+(i,j)$. Therefore, $\Gamma(i,j)$ is different in different image locations; however, δ_k is known for the group G_k defined by image key. This is an additional parameter to be stored along with the image key and could be generated in $O(p)$ time for p pixel locations.

As a consequence, steps 1(c) and 1(d) of Algorithm 1 are modified as follows. Corresponding to the buyer key B, if the bit value $B_k = 0$, then for the group G_k we use $I_w(i,j) = I(i,j) - \Gamma(i,j)$; otherwise for $B_k = 1$, we take $I_w(i,j) = I(i,j)+\Gamma(i,j)$. Thus it is not possible for the collusion attackers to decide on the exact value of each pixel. In such a case the pruning step 2 of Algorithm 3 needs to be modified accordingly.

Modification 2: To add further robustness to the process, we make the image key buyer specific. Therefore, for every multimedia object sold, an image key and a buyer key need to be stored. So, for a specific buyer u of a particular image I, we have image key K_{Iu} and buyer key B_{Iu} so that during retrieval, a bit pattern will be generated for all image keys specific to image I. The owner of the buyer key closest to the retrieved bit pattern is the legal owner of the object. The retrieval method as in Algorithm 3 is modified as follows.

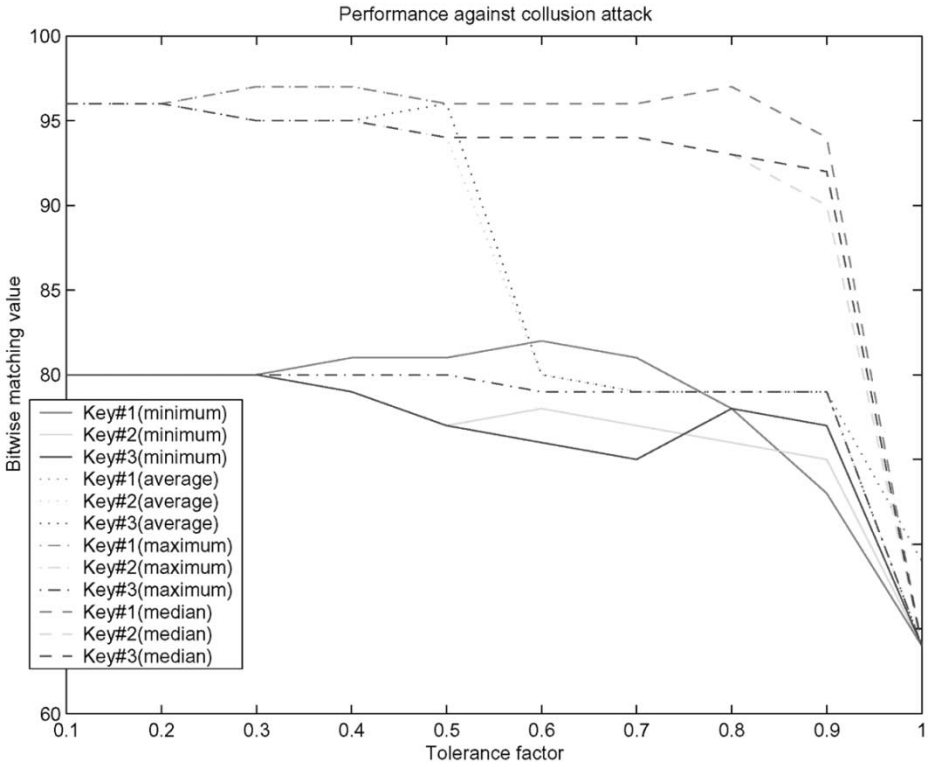


Fig. 12. Performance against collusion attack where three buyers colluded together resulting in four filtered images.

Algorithm 4

1) For $0 \leq i \leq 2^a - 1,\ 0 \leq j \leq 2^b - 1,$
   a) If $I_{w\#}(i,j) > I(i,j) + \Gamma(i,j)$ then $I_{w\#}(i,j) = I(i,j) + \Gamma(i,j)$.
   b) If $I_{w\#}(i,j) < I(i,j) - \Gamma(i,j)$ then $I_{w\#}(i,j) = I(i,j) - \Gamma(i,j)$.
2) For $0 \leq l < \Psi, \Psi$: number of image keys
3) For $0 \leq k \leq 2^n - 1$
   a) Initialize values $\sigma_{Ik} = 0$.
   b) Initialize bit values $q_{Ik} = 0$.
4) For $0 \leq i \leq 2^a - 1,\ 0 \leq j \leq 2^b - 1,$
   If $\langle i,j \rangle$ belongs to the group G_{Ik}, then $\sigma_{Ik} = \sigma_{Ik} + I_{w\#}(i,j) - I(i,j)$
5) For $0 \leq k \leq 2^n - 1$
   If $\sigma_{Ik} < 0$, then $q_{Ik} = 0$; else $q_{Ik} = 1$
6) Compute the code word q' closest to q_I.
7) Report q' as B.

In this context, it is of interest to investigate whether the buyer specific image key helps in identifying one or more of the attackers who participate in collusion attack. Four watermarked copies of the image are taken. Naturally, they have four different buyer and image keys. Intensity averaging of these images at every pixel location creates a forged image. The forged image is then subjected to buyer key retrieval following Algorithm 4. The bit wise matching value between four separate buyer keys with respect to retrieved bit pattern is given in Table VIII. Note that the retrieved bit pattern when checked with Image key #1 specific for Buyer #1 has bit wise matching value 123 with respect to Buyer key #1. This is far greater than the 75% required for exact identification of the buyer involved in the collusion attack. At the same time, the bit wise matching values of this pattern with respect to other buyer keys are far less than 75%. This indicates that no owners will be falsely implicated when watermarks are retrieved. Similar results are obtained when checked with the remaining image keys. The bit wise matching value of the retrieved pattern and a Buyer key (#5) for a buyer who has not colluded in the attacking process is also shown in Table VIII. In this case, the Buyer key #5 is equidistant from all the buyer keys participated in the collusion attack and the matching value is 50%, far less than the threshold 75% required for identification. This again shows that a buyer who has not participated in the collusion attack will be not be incorrectly implicated in a collusion.

As mentioned earlier, following [5], it is possible to design collusion attack that can defeat the proposed variation of our watermarking process. Given a buyer specific image key and a variation of $\beta(i,j)$ for every pixel location, it is difficult to assess the number of adversaries required to defeat the process. For a given image key that divides the image into m groups, the attackers need to defeat our algorithm in more than $m/4$ groups. In other words, they have to assess the total variation in pixel values for each of the $m/4$ groups. For every pixel, given $\mu^-(i,j) \leq \beta(i,j) \leq \mu^+(i,j)$, there are $(2|\mu(i,j)| + 1)$ possible ways the pixel values are changed. So, in a straightforward manner, the attackers need to tackle more than $(2|\mu(i,j)| + 1) \times 2^{a+b-n} \times m/4$ possible pixel manipulations, which is computationally prohibitive. From the experimentation using cropping attacks in Section IV, roughly, if more than $\sim 60\%$ of the image regions are randomly replaced by the colluders with pixels having

$\beta(i,j) = 0$ (that is the original image pixels), the buyer key cannot be recovered.

To show the advantage of error correction while combating a collusion attack, we have performed an experiment in which three buyers colluded together to generate an attacked image. Three separate buyer keys are used to generate three watermarked images. The minimum, maximum, average and median of corresponding pixel values are calculated using three watermarked images [2]. This results in four different attacked images. If buyer keys can be extracted, the owner can identify the buyers involved in the collusion process. The bitwise matching values are shown in the graph of Fig. 12. While the maximum bitwise matching values of 97–99 bits (of 128) are obtained in case of average, median and maximum version of the attacked image, maximum bit wise matching of 80 bits is obtained for minimum version of the attacked image. Therefore, if we could select a 128 bit length buyer key with minimum Hamming distance between them being 100, we could always correct the retrieved buyer key in the case where at least $(128 - (100/2)) = 78$ bitwise matches are found between the retrieved bit pattern and the buyer key in the database.

VI. Conclusion

We have proposed a novel watermarking technique that survives attacks both in frequency and spatial domains. The strength of the algorithm is demonstrated through survival of the proposed watermark after Stirmark and JPEG compression attacks. The motivation of our watermarking scheme is twofold: first, we have assumed that the attacker knows the entire watermarking process, and second, the watermarking process could be individualized and linked with respect to the specific owner of the multimedia object. As explained in the methodology section, the watermark retrieval is based on the secure image key and the original image. Since the retrieval is basically the identification of the buyer key, not only the authenticity can be proved but also the trail of forging can be identified through the buyer key.

As shown in [5], the survival of a watermark under a full scale collusion attack is still an open problem. Our method shows a significant number of owners need to come together in order to have a successful collusion attack. The proposed technique is computationally attractive and has the potential for improvement. The scheme is also suitable for watermarking in the frequency domain. We are currently investigating watermarking application in real time audio and video sequences.

Acknowledgment

The authors would like to thank T. K. Das of Indian Statistical Institute for excellent programming support and providing important suggestions that has improved the technical quality of the paper.

References

[1] C. Cachin, "An information-theoretic model for steganography," presented at the 2nd Workshop on Information Hiding, Portland, OR, Apr. 1998.

[2] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.

[3] I. J. Cox and M. L. Miller, "A review of watermaking and the importance of perceptual modeling," presented at the Electronic Imaging '97, Feb. 1997.

[4] S. Craver, B. Yeo, and M. Yeung, "Technical trials and legal tribulations," *Commun. ACM*, vol. 41, no. 7, 1998.

[5] F. Ergun, J. Kilian, and R. Kumar, "A note on the limits of collusion-resistant watermarks," in *Eurocrypt 1999*, vol. 1592, Lecture Notes in Computer Science, 1999, pp. 140–149.

[6] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1988, pp. 16–18.

[7] J. R. Hernansez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermaking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Processing*, vol. 9, pp. 55–68, Jan. 2000.

[8] C. Honsinger and M. Rabbani, "Data Embedding Using Phase Dispersion, Kodak Tech. Rep.," Imaging Science Div., Rochester, NY.

[9] N. F. Johnson, Z. Duric, and S. Jajodia, *Information Hiding: Steganography and Watermarking—Attacks and Countermeasures*. Norwell, MA: Kluwer, 2000.

[10] G. C. Langelaar and R. L. Lagendijk, "Optimal differential energy watermarking of DCT encoded images and video," *IEEE Trans. Image Processing*, vol. 10, pp. 148–158, Jan. 2001.

[11] C. S. Lu, S.-K. Huang, C.-J. Sze, and H.-Y. Liao, "A new watermarking technique for multimedia protection," in *Multimedia Image and Video Processing*, L. Guan, S.-Y. Kung, and J. Larsen, Eds. Boca Raton, FL: CRC, 2001, pp. 507–530.

[12] F. J. MacWilliams and N. Sloane, *The Theory of Error-Correcting Codes—Part I*. Amsterdam, The Netherlands: North Holland, 1977.

[13] P. Meerwald, "Digital Watermatrking in the Wavelet Transform Domain," Master's, Dept. Sci. Comput., Univ. Salzburg, Austria, 2001.

[14] F. Petitcolas, R. J. Anderson, M. G. Kuhn, and D. Aucsmith, "Attacks on copyright marking systems," in *Proc. 2nd Workshop on Information Hiding*, Portland, OR, April 1998, pp. 218–238.

[15] J. O. Ruanaidh, H. Petersen, A. Herrigel, S. Pereira, and T. Pun, "Cryptographic copyright protection for digital images based on watermarking techniques," *Lecture Notes Theoret. Comput. Sci.*, vol. 226, pp. 117–142, 1999.

[16] S. Voloshynovskiy, S. Pereira, V. Iquise, and T. Pun, "Attack modeling: Toward a second generation watermarking benchmark," *Signal Process., Special Issue on Information Theoretic Issues in Digital Watermarking*, vol. 81, no. 6, pp. 1177–1198, June 2001.

[17] M. M. Yeung, "Digital watermarking," *Commun. ACM*, vol. 41, no. 7, 1998.

Dipti Prasad Mukherjee (M'01) received the B.E. degree from Jadavpur University, Calcutta, India in 1985, the M.S. degree from the University of Saskatchewan, Saskatoon, SK, Canada in 1989, and the Ph.D. degree from the Indian Statistical Institute (ISI), Calcutta, in 1996.

He is currently a faculty member with the Electronics and Communication Sciences Unit, ISI. He was a Visiting Assistant Professor at Oklahoma State University, Stillwater, in 1998–1999 and Research Scientist in the Electrical and Computer Engineering Department, University of Virginia, Charlottesville, in 2002. He was a UNDP fellow to the Robotics Research Group, University of Oxford, U.K., in 1992. His research interests are in the areas of computer vision and graphics. He has published more than 25 peer-reviewed journal papers and is the author of a textbook on computer graphics and multimedia.

Dr. Mukherjee was the recipient of UNESCO-CIMPA fellowships to INRIA, France, in 1991, 1993 and 1995 and fellowships to ICTP, Trieste, Italy, in 2000.

Subhamoy Maitra received the B.Eng. degree in electronics and telecommunication engineering from Jadavpur University, Calcutta, India, in 1992, and the M.Tech. degree in computer science in 1996 from the Indian Statistical Institute (ISI), Calcutta, and the Ph.D. degree from ISI in 2001.

He is currently with the faculty at ISI. His research interest is in cryptology and digital watermarking.

Scott T. Acton (M'93–SM'99) received the B.S. degree in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, in 1988 as a Virginia Scholar. He received the M.S. degree and the Ph.D. degrees in electrical engineering from the University of Texas at Austin in 1990 and 1993, respectively.

He is currently an Associate Professor, Department of Electrical and Computer Engineering, University of Virginia. He has worked in industry for AT&T, the MITRE Corporation, and Motorola, Inc., and in academia for Oklahoma State University, Stillwater.

Dr. Acton is the winner of the 1996 Eta Kappa Nu Outstanding Young Electrical Engineer Award, a national award given annually since 1936. He also received the 1997 Halliburton Outstanding Young Faculty Award. He has served as Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and is currently serving as Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. His research interests include biomedical image analysis, multiscale signal representations, diffusion algorithms, active contours, video tracking, image morphology, image segmentation, and content-based image retrieval.