Data Hiding in Image and Video: Part II—Designs and Applications

Min Wu, *Member, IEEE*, Heather Yu, *Associate Member, IEEE*, and Bede Liu, *Fellow, IEEE*

***Abstract*—This paper applies the solutions to the fundamental issues addressed in Part I to specific design problems of embedding data in image and video. We apply multilevel embedding to allow the amount of embedded information that can be reliably extracted to be adaptive with respect to the actual noise conditions. When extending the multilevel embedding to video, we propose strategies for handling uneven embedding capacity from region to region within a frame as well as from frame to frame. We also embed control information to facilitate the accurate extraction of the user data payload and to combat such distortions as frame jitter. The proposed algorithm can be used for a variety of applications such as copy control, access control, robust annotation, and content-based authentication.**

***Index Terms*—Data hiding, digital watermarking, multilevel embedding, video data hiding.**

I. Introduction

In Part I [1], we have addressed a few fundamental issues of data hiding in image and video. We have proposed general solutions, including how to embed multiple bits, how to handle uneven embedding capacity, and how to allow the number of reliably extractable bits to be adaptable to the actual noise condition. Here in Part-II, we apply the solutions to specific design problems and present details of embedding data in image and video.

In Section II, we embed data in images at two levels, each of which is designed for different robustness. This approach allows for graceful decaying of extractable information as noise gets stronger. In Section III, we extend the multilevel embedding to video, for which difficulty arises because the embedding capacity varies from region to region within a frame as well as from frame to frame. We embed control information to facilitate the extraction of the user data payload and to combat such distortions as frame jitter.

The designs presented in this paper can be used as building blocks for such applications as copy control, access control, robust annotation, and content-based authentication. Comprehensive protection from malicious attacks that make watermarks undetectable would require both technical and business approaches, such as a well-determined business and pricing model. Our design objective here focuses on surviving common processing in transcoding and scalable/progressive transmission, such as compression with different ratio and frame rate conversion for video.

Manuscript received February 4, 2002; revised November 22, 2002. This work was supported in part by Panasonic Information and Networking Laboratory, by a R&D Excellence Grant from the State of New Jersey, and by the National Science Foundation CAREER Award CCR-0133704. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bruno Carpentieri.

M. Wu is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: minwu@eng.umd.edu).

H. Yu is with Panasonic Information and Networking Laboratories (PINTL), Princeton, NJ 08540 USA (e-mail: heathery@research.panasonic.edu).

B. Liu is with Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: liu@ee.princeton.edu).

Digital Object Identifier 10.1109/TIP.2003.810589

II. Multilevel Data Hiding in Grayscale Image

In this section, we present a two-level data hiding using the two types of embedding mechanisms discussed in Part-I. The basis of this section is Fig. 5 of Part I, which demonstrates that by combining several embedding levels, the number of bits that can be reliably extracted will decay gracefully as the actual noise gets stronger.

We focus here on how to convey several sets of data with different robustness, and depending on the applications, the data in each set could be either identical or be different [2], [3]. We consider that the amount of data in each set is nontrivial. The case of using one embedding level to convey a small amount of side information to facilitate the extraction of the main payload will be discussed in Section III-C.

For simplicity, we study the problem of multilevel data hiding in grayscale images. Extension to color images is straightforward. The embedding domain we have chosen is the 8×8 block DCT coefficients. This domain is compatible with commonly used image and video compression standards, making it easier to perform compressed domain embedding and to apply known results such as human visual models for JPEG compression [4], [5]. It also allows for fine tuning of the watermark strength for each local region to achieve good tradeoff between imperceptibility and robustness.

We use nonoverlapped spectrum segments for multiple-level embedding to avoid interference among different levels, although overlapped embedding can also be used[1]. A key issue in nonoverlapped embedding is to determine what part of the host signal to be used for each embedding level. The following analysis on the performance of noncoherent detection of Type-I spread spectrum embedding provides a guideline to the partitioning of host signal spectrum for two-level data hiding.

[1]Overlapped embedding is similar to the embedding of two or more watermarks successively into a host signal to simultaneously achieve multiple goals [6]–[8]. For example, a robust watermark and a fragile watermark can be added to an image for ownership protection and tampering detection, respectively.

1057-7149/03$17.00 © 2003 IEEE

A. Spectrum Partition

Using Type-I embedding to hide one-bit information in the host signal under additive noise can be formulated as a hypothesis testing problem

$$\begin{cases} H_0 : y_i = -s_i + d_i \ (i = 1, \ldots, n), & \text{if } b = -1 \\ H_1 : y_i = +s_i + d_i \ (i = 1, \ldots, n), & \text{if } b = +1 \end{cases} \tag{1}$$

where the watermark $\{s_1, \ldots, s_n\}$ is an n-sample known sequence, b is a bit to be embedded and is equally likely to be "+1" or "−1", and d_i represents the total noise and interference. Under the assumption that d_i is i.i.d. Gaussian with density $\mathcal{N}(0, \sigma_d^2)$, the optimal detection statistic is a correlator

$$T = \frac{\underline{y}^T \underline{s}}{\sqrt{\sigma_d^2 \cdot \|\underline{s}\|^2}} \tag{2}$$

which is Gaussian distributed with unit variance and mean

$$E(T) = b \cdot \sqrt{n \cdot \frac{\left(\frac{1}{n}\|\underline{s}\|^2\right)}{\sigma_d^2}}. \tag{3}$$

Setting the threshold to zero gives minimum probability of error $\mathcal{Q}(E(T))$, where $\mathcal{Q}(x)$ is the probability $P(X > x)$ of a Gaussian random variable $X \sim \mathcal{N}(0, 1)$.

Under noncoherent detection, d_i consists of the interference from host media and the noise due to processing and attack. The high power of host media contributes to a large σ_d^2 value, increasing the probability of detection error. A popular approach to reduce the error probability is to only watermark mid-band coefficients [9] and to leave the low band and high band unchanged. It is based on the observation that the low band coefficients generally have much higher power than those in mid-band, and that the high band coefficients are vulnerable to processing and attacks. Also, modification of low band coefficients may have a higher impact on watermark perceptibility.

It is possible, however, to embed in the low band, provided perceptual model is used and the effect of large values on σ_d^2 is taken into account. The test statistic T of (2) is optimal if the noise $\{d_i\}$ is i.i.d. Gaussian, which often does not hold in practice. A better yet simple assumption is that $\{d_i\}$ is independent Gaussian, but with different variance for different frequency bands. The optimal detector is then a correlator preceded by normalizing the observations with their corresponding standard deviations σ_{d_i}, which gives more weight to less noisy components. The test statistic then becomes

$$T' = \frac{\sum_{i=1}^{n} \frac{y_i \cdot s_i}{\sigma_{d_i}^2}}{\sqrt{\sum_{i=1}^{n} \frac{{s_i}^2}{\sigma_{d_i}^2}}}. \tag{4}$$

Thus, it is possible to embed data in all bands, although contributions from those noisy bands are limited. One can also use a more general Gaussian noise model in which the components of the host media and/or the noise may be dependent. In this case, both whitening and normalization are performed before applying the minimum Euclidean distance detector or maximum correlation detector [10].

1) Verification Through Experiments: The above analysis is verified experimentally using 114 photographic images and the block-DCT spread spectrum algorithm proposed by Podilchuk-Zeng [11]. For detection, the *q-statistic* proposed by Zeng-Liu [12] is used. We denote by q' and q the detection statistic with and without the weighting based on an estimation of the total noise in each band, respectively. That is

$$q = \frac{M_Z}{\sqrt{\frac{V_Z}{n}}} \tag{5}$$

$$q' = \frac{M_{Z'}}{\sqrt{\frac{V_{Z'}}{n}}} \tag{6}$$

where

$$Z_i = y_i \cdot s_i, \quad Z'_i = y_i \cdot \frac{s_i}{\gamma_i}$$

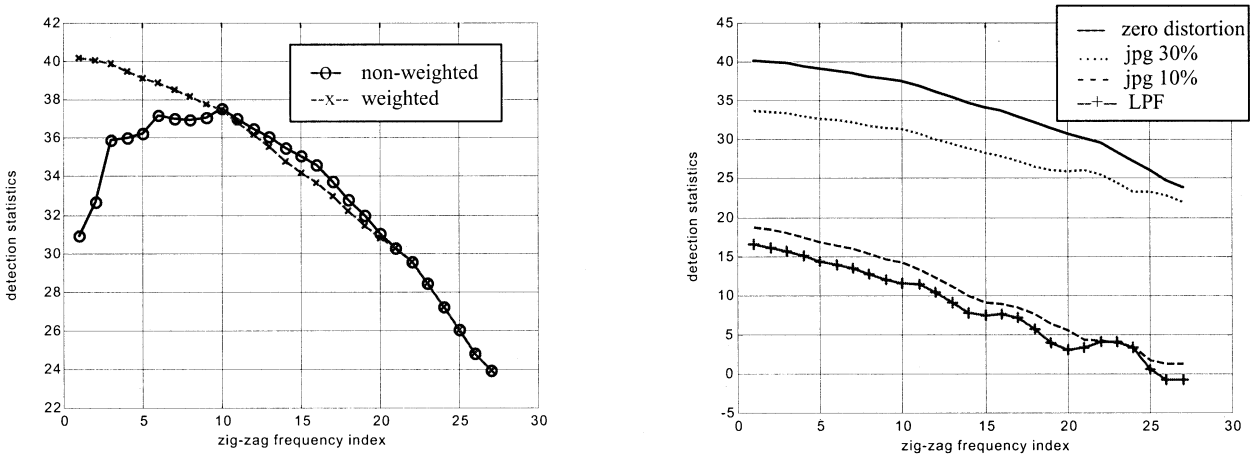
$$M_Z = \frac{1}{n}\sum_{i=1}^{n} Z_i, \quad M_{Z'} = \frac{1}{n}\sum_{i=1}^{n} Z'_i$$

$$V_Z = \frac{\sum_{i=1}^{n} (Z_i - M_Z)^2}{n-1}, V_{Z'} = \frac{\sum_{i=1}^{n} (Z'_i - M_{Z'})^2}{n-1}.$$

The weight $\{\gamma_i\}$ reflects the impact of the noise variance term in (4). The q statistic of (5) is a correlation with variance normalized to 1 without explicitly estimating the variance of noise and interference σ_d^2.

The noise variance is not easy to estimate accurately because the precise power of the host signal is unknown in noncoherent detection, and the variance of processing noise is highly dependent on the distortion or attack applied to the signal. To overcome these difficulties, an estimate of host signal power can be made using the current test image. A set of known signal can be added to predetermined locations of the host signal, serving as a set of training data to facilitate the noise estimation [13]. The $\{\gamma_i\}$ in our experiment is based on the variance of host signal and potential processing noise of the frequency band of y_i. They are empirically determined using a collection of natural images.

Using q and q' as detection statistics, each of the above-mentioned 114 natural images is tested using three different spread spectrum watermarks. For each watermark and each image, the block DCT coefficients are ordered in the familiar zig-zag manner (Fig. 2). We then vary the frequency beyond which a watermark is inserted. The q and q' values are computed under several distortion conditions including no distortion, JPEG with different quality factors, and low pass filtering. For each image, we also normalize q and q' with respect to the number of *embeddable* coefficients that can be watermarked without introducing perceptual distortion. The average normalized q and q' are shown in Fig. 1, where the horizontal axis is the zig-zag ordered frequency band beyond which data is embedded. It can be seen that q is maximum when the band from which the embedding starts is around 6 to 11, and it falls off when either more or less number of frequency bands are involved. It is also seen that q' is larger than q, hence q' gives a smaller probability of error. In addition, q' is monotonically decreasing when fewer bands are used in embedding, but the decrease is insignificant



(a) (b)

Fig. 1. Average detection statistics: (a) detection statistics of nonweighted correlator q (circles) and of weighted correlator q' (crosses) with no additional distortion and (b) detection statistics of weighted correlator q' under four different distortions. The x-axis in both plots indicates the frequency band in a zigzag order from which watermark starts to be put in.

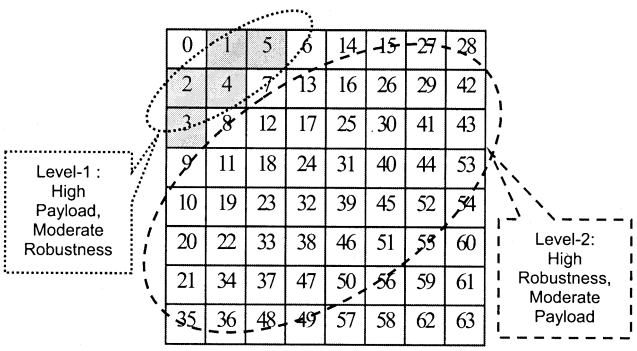


Fig. 2. Spectrum partition of two-level data hiding in block-DCT domain.

when leaving out the first five lowest bands from embedding. These observations are consistent with our analysis.

The above study suggests that for a two-level embedding system, one should apply the Type-I spread spectrum embedding to mid-band coefficients for high robustness at a cost of total payload, and apply Type-II enforcement embedding to low-band for high payload with moderate robustness. Such a multilevel embedding approach would allow the hiding of many bits and decode them successfully when image experiences little or moderate distortion. When an image is distorted significantly, this approach can still reliably extract those bits that have been embedded robustly.

B. System Design

Shown in Fig. 3 are block diagrams of two-level data hiding in image. The first level uses odd-even embedding in the low band, which are the first two diagonal lines of AC coefficients (Fig. 2). The embedding is done with quantization step sizes $\{Q_i\}$ to enhance robustness. That is, a watermarked coefficient v'_i is obtained from the original coefficient v_i of the host signal using

$$v'_i = \left(\text{round}\left[\frac{v_i}{Q_i}\right] + \delta\right) \cdot Q_i. \tag{7}$$

δ is determined by

$$\delta = \begin{cases} 0, & \text{if mod}\left(\text{round}\left[\frac{v_i}{Q_i}\right], 2\right) = b_i; \\ \text{sgn}\left(\frac{v_i}{Q_i} - \text{round}\left[\frac{v_i}{Q_i}\right]\right), & \text{otherwise} \end{cases} \tag{8}$$

where $b_i \in \{0, 1\}$ is the bit to be embedded, and $\text{sgn}() \in \{-1, +1\}$ is the signum function. We use the quantization step sizes that are equivalent to the standard JPEG quantization table of quality factor 50% [14]. If the changes from v_i to v'_i is larger than the just-noticeable-difference (JND), that coefficient is regarded as unembeddable, and no changes are made to it. The human visual model used here is refined from the frequency-masking model by Podilchuk-Zeng [11] to reduce ringing artifacts [15]. Local image statistics are used to distinguish texture and edge blocks and to attenuate the JND of edge blocks.

The second set of data is embedded in mid-band using Type-I additive spread spectrum technique. Antipodal modulation is used by adding or subtracting a spread spectrum signal, $\{w_i\}$, to represent one bit

$$v'_i = v_i + b' \cdot \alpha_i \cdot w_i, \quad i = 1, \ldots, n \tag{9}$$

where $\{v_i\}$ are the original coefficients, $\{v'_i\}$ are the marked coefficients, and $b' \in \{-1, +1\}$ is the antipodal mapping from b, the bit to be embedded. The watermark strength, $\{\alpha_i\}$, is adjusted by JND.

TDM-type multiplexing/modulation (Part I, Sec. IV) is used at both levels. The bits are embedded in nonoverlapped regions. For Level-1, the low band coefficients of all blocks are divided into several distinct sets, and in each set a bit is embedded using odd-even enforcement on all coefficients. The detector determines a bit by majority voting over the extracted values from those coefficients. For each bit embedded in Level-2 (high robustness), we partition a spreading sequence into nonoverlapped segments and assign one segment to that bit. To overcome uneven embedding capacity of TDM, coefficients for each of the

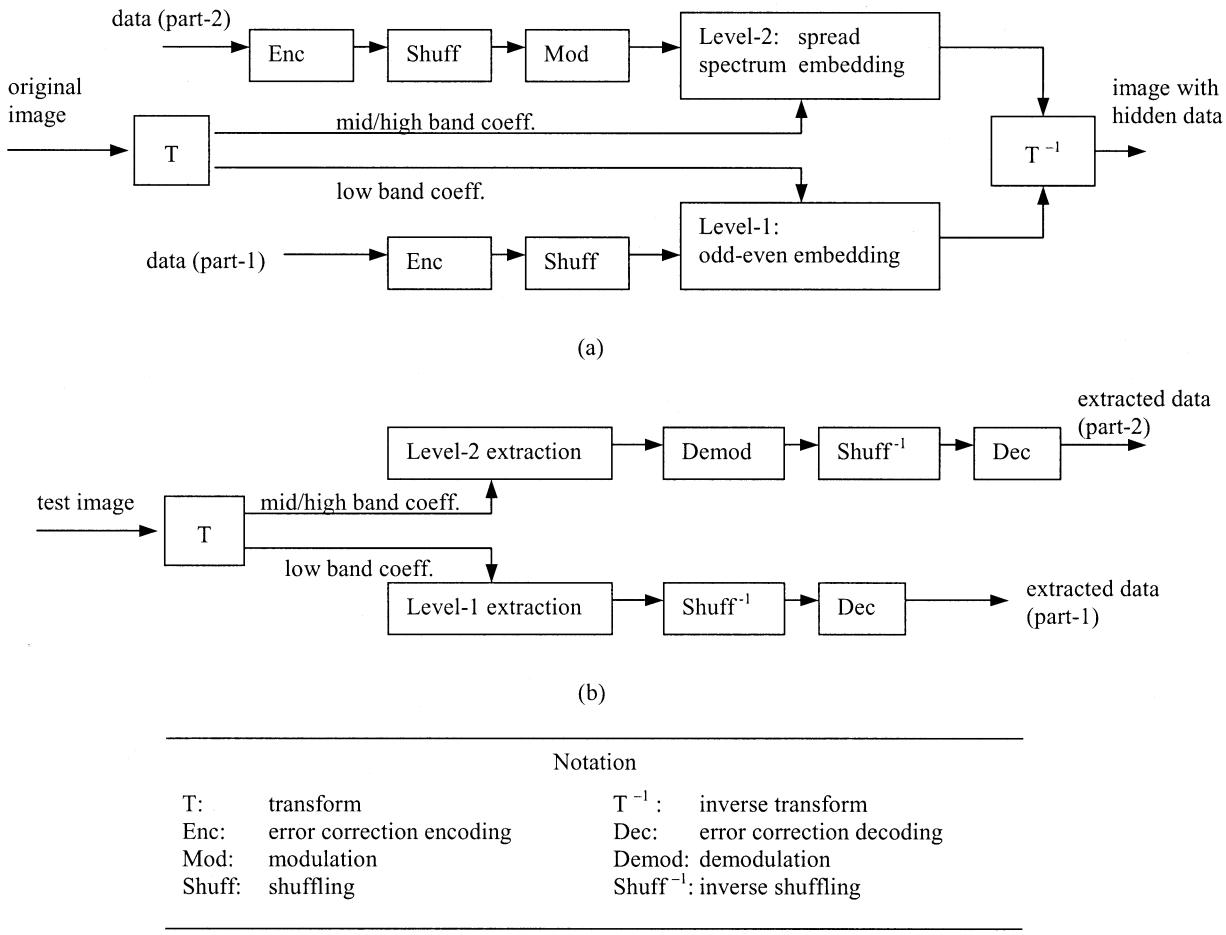


Fig. 3. Block diagram of two-level data hiding for images: (a) embedding process and (b) extraction process.

two embedding levels are shuffled and the embedding is performed in shuffled domain (Part I, Sec. V). An inverse shuffling and an inverse DCT transform are then applied to obtain watermarked image. The data embedded in each of the two levels can be further encoded using error correction codes.

This design of a two-level data hiding system serves as a proof-of-concept of our proposed multilevel embedding (Part I, Sec. III). Other embedding schemes with different payload-robustness settings can also be incorporated to meet the needs of different applications.

C. Experimental Results

We apply the proposed two-level data hiding scheme to the 512×512 Lenna image shown in Fig. 4(a). The watermarked image, Fig. 4(b), has a PSNR of 42.5dB with respect to the original unmarked image. Incorporating BCH error correction coding and shuffling, we embed a 32×32 binary pattern of PINTL-Matsusita logo in low band, which can be extracted accurately when the image experiences JPEG compression of quality factor 45% or higher. We also use spread spectrum technique to embed the ASCII code of a character string "PINTL" in mid-band, which can be extracted without error when the image is blurred or JPEG compressed with quality factor as low as 20%. The embedding rate can be higher for images that contain larger textured region. For example, we can embed a longer string of "Panasonic Tech." and the 32×32 PINTL-Matsusita pattern in the Baboon image, as shown in Fig. 5. Using our refined human visual model, the marked image has no visible artifacts and has a PSNR of 33.6dB with respect to the original image. The lower PSNR of the Baboon image than that of the Lenna image is a result of the Baboon image having more textured regions.

III. Multilevel Data Hiding in Video

In this section, we extend multilevel embedding from image to video, guided by the general results from Part-I. The issues involved in data hiding in video, besides the large data volume and high computation complexity, are the selection of an appropriate embedding domain and the handling of uneven embedding capacity.

A. Embedding Domain

Consecutive frames in a video look similar except those at scene changes or with fast motion. Because of this, it is possible to add or drop some frames, or switch the order of adjacent frames, without causing much noticeable artifacts. In addition, new frames may be generated from a few similar frames and inserted to the sequence or replace some original frames. If different data are embedded in each frame of the original video and several watermarked frames are used to generate a new frame, the embedded data may not be easily detectable from

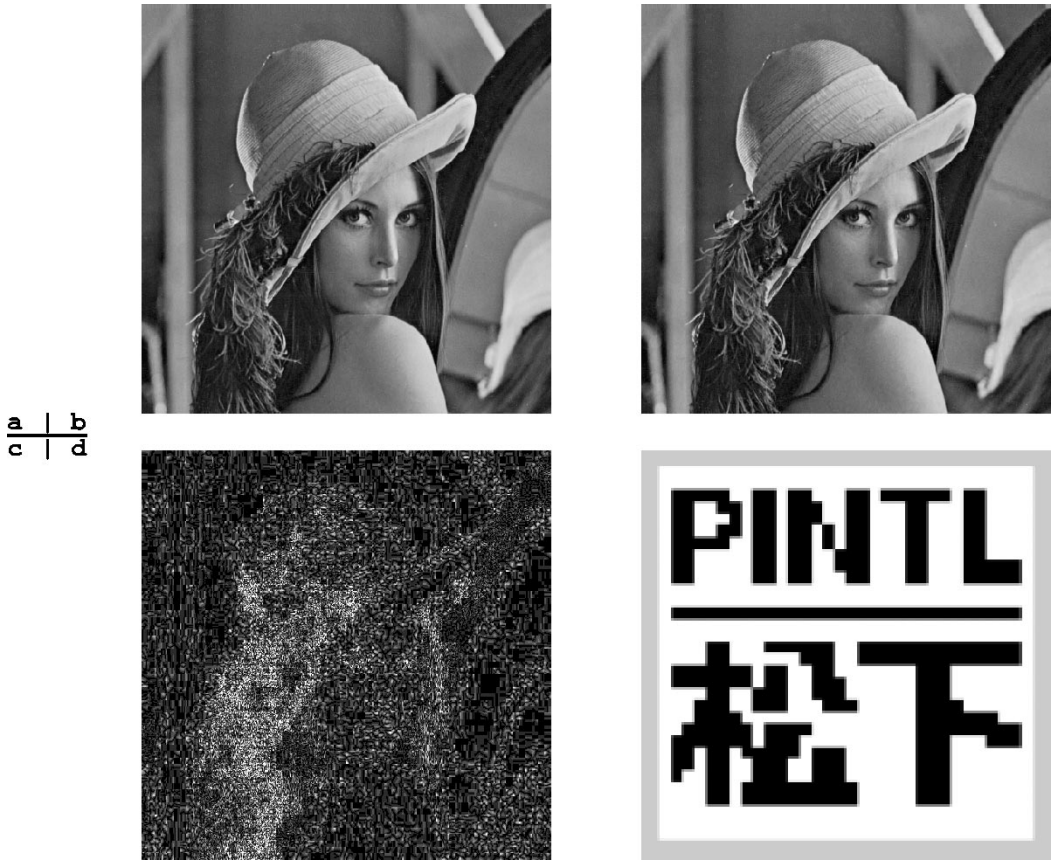


Fig. 4. Multilevel data hiding for Lenna image (512 × 512). (a) original image; (b) watermarked image; (c) amplified difference (× 5) between (b) and (a) with black denoting zero difference; and (d) extracted 32 × 32 PINTL-Matsusita logo from the low band.

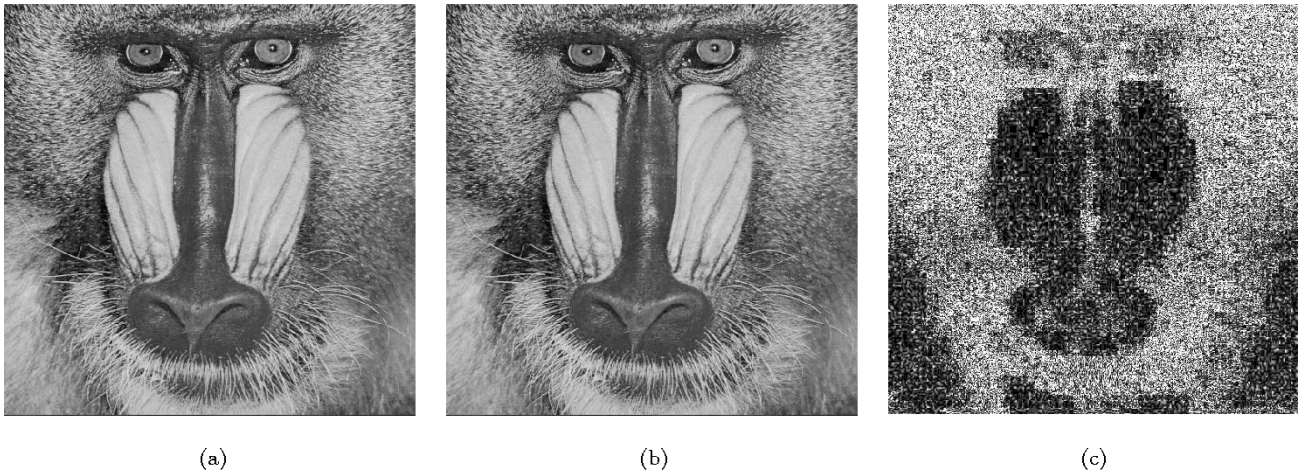


Fig. 5. Multilevel data hiding for Baboon image (512 × 512). (a) original image; (b) watermarked image; (c) amplified difference (× 5) between (b) and (a) with black denoting zero difference.

these new frames. This is known as collusion attack [16]. Since such manipulations can arise from common processing involved in format conversion and transcoding [17] or from malicious attacks, these possibilities must be considered in the design of robust data hiding for video. Adding redundancy and searching for frame-jitter invariant domain are common ways to handle these attacks. We focused on the redundancy approach because of its effectiveness and computational simplicity.

We handle frame jitter by first partitioning the video into temporal segments, and each consists of similar consecutive frames.

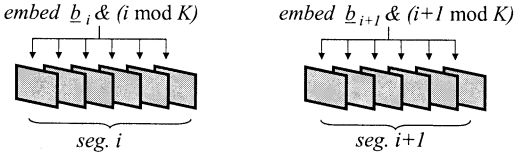


Fig. 6. Illustration of methods for handling frame jittering.

We embed the same data in every frame of a segment, as illustrated in Fig. 6. The temporal partition should be content based,

TABLE I
ADAPTIVE EMBEDDING RATE FOR A VIDEO FRAME

Estimated Achievable Payload $\hat{C}$	*Embedding Rate for User Data*	*Corresponding Control Data*
$\hat{C} \leq \tau_1$	Zero Rate - no user bits are embedded	add $+\underline{c}_2$
$\tau_1 < \hat{C} < \tau_2$	Low Rate - hide a small, predefined number of bits	add $+\underline{c}_1$
$\hat{C} \geq \tau_2$	Higher Rate - the # of bits embedded is determined by $\hat{C}$	add $-\underline{c}_1$, and use a few spread spectrum sequences to convey the # of bits embedded

since frames before and after a scene change or a big change due to motion can have significantly different embedding capability. Thus the lengths of segments may not be always uniform, although between scene changes we can simply partition the video into segments of equal length. Repetition alone is neither able to handle segments of unequal length, nor is it effective to combat frame reordering, frame insertion, and frame dropping of larger units. Our approach is to embed the same user data as well as a shortened version of segment index in each frame. This *frame sync* index can assist in detecting and locating frame jittering, and is part of the *control bits* that will be addressed in detail in Section III-C. This approach is effective against frame dropping that involves a small number of isolated frames. When this approach is used in conjunction with other redundancy approaches such as repeatedly embedding the same data in separate parts of a long video, the robustness against frame jittering can be further enhanced. Embedding the same data in a segment of frames also provides redundancy to combat the noise from additional processing or attacks. Extraction can be done via weighted majority voting with larger weights assigned to the frames experiencing less distortion.

It should be noted that repeatedly embedding the same data in several consecutive frames is not equivalent to embedding data in the corresponding averaged frame. This is because the embedding operation is nonlinear in general. For Type-II enforcement embedding, the relations such as the odd-even parity enforced on an averaged frame often does not hold in each individual frame or the average of a subset of these frames, hence does not survive frame jitter well. And for Type-I additive embedding, the same JND model gives significantly different result in determining what DCT coefficients are embeddable. Since averaging several consecutive frames is equivalent to temporal low pass filtering, less DCT coefficients in the middle band of an averaged frame will be deemed embeddable than those of the original frames.

B. Variable Embedding Rate (VER) Versus Constant Embedding Rate (CER)

For video, the uneven embedding capacity arises both from region to region within a frame and from frame to frame. As discussed in Part-I, VER requires a nontrivial amount of side information but could provide higher overall embedding payload, and CER requires only a small amount of one-time side information but may be wasteful in total embedding capacity. Here we shall combine VER and CER as the follows. The intra-frame unevenness is handled using CER and shuffling, and VER is used for inter-frame unevenness with the help of additional side information. An equal number of bits are embedded in each group of shuffled coefficients within a frame. The group size, or equivalently, the number of bits embedded in each frame, is different from frame to frame and depends on an estimated achievable payload discussed below. The overhead is thus relatively small compared to the total number of bits that can be embedded in most frames.

The number of bits that can be embedded in each frame may vary from very few bits for smooth frames to dozens or even hundreds bits for frames containing large regions of details and textures. Variable length codes can be used to represent this side information, with shorter codes assigned to those frames that can have only a small number of bits embedded. For each video segment, we estimate the achievable embedding payload $\hat{C}$ per frame based on the energy of DCT coefficients, the number of embeddable DCT coefficients, and the detection statistic of an embedding trial that hides only a single spread spectrum watermark in a video frame. We also set two thresholds τ_1 and τ_2. If $\hat{C} \leq \tau_1$, we embed no user data. If $\tau_1 < \hat{C} < \tau_2$, a predefined number of user bits are embedded. If $\hat{C} \geq \tau_2$, we embed user data at a higher rate determined by $\hat{C}$. Table I summarizes the adaptive determination of embedding rate. We use spread spectrum sequences $+\underline{c}_2$, $+\underline{c}_1$, and $-\underline{c}_1$ to signal the aforementioned three cases, respectively. In the case of $\hat{C} \geq \tau_2$, we also use orthogonal modulation via several other spread spectrum sequences to convey the number of embedded bits. To reduce the overhead for conveying this side information, we limit the number of embedded bits to one of a pre-determined finite set (e.g., $\{16, 32, 48, 64, \ldots\}$), which can be determined empirically using training video clips. All these are part of the control data that need to be conveyed to facilitate the extraction of user payload data. We will discuss more about embedding control data in the next subsection.

The estimated achievable payload $\hat{C}$ is determined as the follows. For Type-I additive spread spectrum embedding, the mean detection statistic $E(T)$ is given by (3) and follows a unit-variance Gaussian distribution. The bit error probability is $\mathcal{Q}(E(T))$. Given the maximum bit error probability $P_e^{(\max)}$ that can be tolerated by the application, a lower bound of mean detection statistic required for each bit is $T_{th} = \mathcal{Q}^{-1}(P_e^{(\max)})$. We denote the detection statistic when all embeddable coefficients are used to carry one information bit as T_0. The estimated number of bits that can be embedded is thus upper bounded by

$$\hat{C} = \left(\frac{T_0}{T_{th}}\right)^2. \tag{10}$$

In our experiments, we set T_{th} to be around 5. Similarly, $\hat{C}$ for Type-II enforcement embedding is estimated based on the number of embeddable coefficients to whom the relations can be enforced.

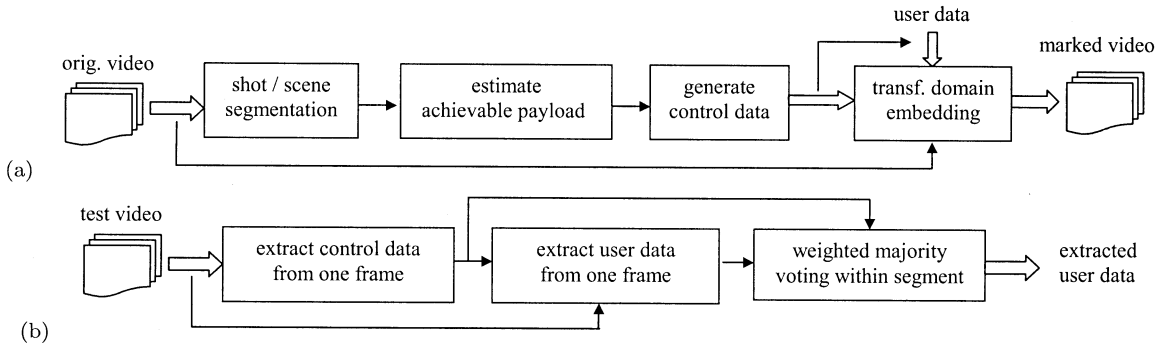


Fig. 7. Block diagram of the proposed video data hiding system: (a) embedding process and (b) detection process.

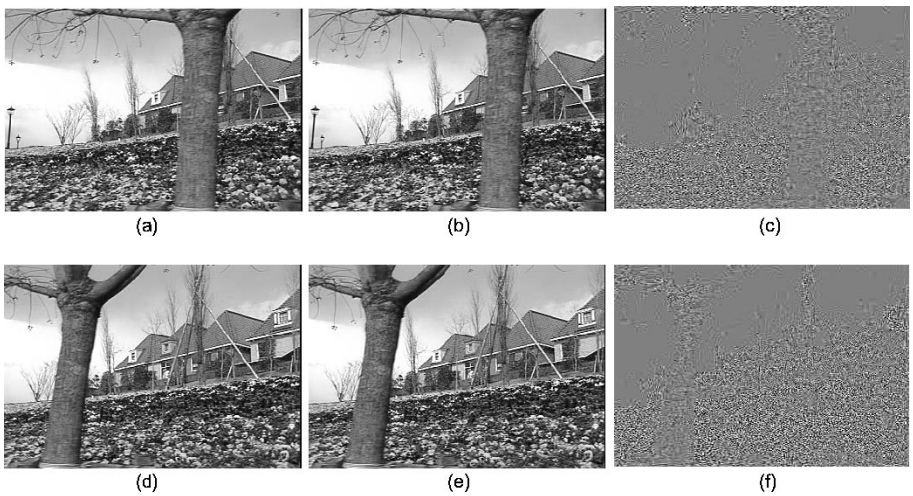


Fig. 8. Multilevel data hiding for the flower garden video sequence: (a) the original 1st frame, (b) the watermarked 1st frame, (c) the amplified difference of (a) and (b); (d) the original 30th frame, (e) the watermarked 30th frame, and (f) the amplified difference of (d) and (e). The video is compressed with MPEG-2 at 4.5Mbps, and the differences are amplified by a factor of 5 with gray denoting zero difference and black/white denoting large difference.

C. Control Data Versus User Data

Additional information, known as *control data*, is often needed to facilitate extraction of *user data* or *user payload*. In our design, the control data include the frame sync index, the number of bits embedded in each frame, and a constant watermark for image registration when the video is subject to geometric distortion [18], [19].

The amount of control data is relatively small when compared to that of user data, but is critical and should be extracted accurately. Thus we use the robust spread spectrum embedding and the energy-efficient orthogonal modulation to embed control information, where the spreading sequence for hiding each control bit is orthogonal with one another and is also orthogonal to those used for user data.

Here we use frame sync as an example to demonstrate the embedding of control data. As introduced in Section III-A, frame sync is a short version of video segment index. Its range is from 0 to $K-1$, i.e., the ith segment is labeled with an index of $\mathrm{mod}(i, K)$. A larger K takes more bits, but gives better tolerance to frame jitter. Experiments have shown that $K = 8$ is a good choice. The video segments are then indexed in a round-robin fashion from 0 to 7 and each index is embedded using the orthogonal modulation discussed in Part-I. So for the sync index j, we embed the j-th sequence of K pre-selected orthogonal random sequences.

User data is embedded in each video frame using the multilevel approach discussed in Section II. TDM with shuffling is applied when hiding multiple bits at the high payload level via odd-even enforcement. For the high-robustness embedding via spread spectrum technique, we combine TDM and orthogonal modulation (Part-I Sec.V) to double the number of embedded bits from using TDM alone. As such, a watermark conveying $2B$ bits is formed by

$$\underline{w} = \sum_{k=1}^{B} b_k \cdot \left[\mathcal{I}(b_{B+k} = 1) \cdot \underline{u}_k^{(1)} + \mathcal{I}(b_{B+k} \neq 1) \cdot \underline{u}_k^{(2)}\right] \tag{11}$$

where $b_i \in \{+1, -1\}$, and $\mathcal{I}(\cdot)$ is an indicator function. We first generate two orthogonal spreading sequences $\{\underline{u}^{(1)}\}$ and $\{\underline{u}^{(2)}\}$, and break each sequence into B nonoverlapped segments (TDM) to form the orthogonal spreading vectors $\{\underline{u}_k^{(1)}\}$ and $\{\underline{u}_k^{(2)}\}$, respectively.

D. Experimental Results

A block diagram of the proposed video data hiding system is shown in Fig. 7. The details of the modules that perform data embedding and extraction within each frame are similar to the multilevel image data hiding in Section II.

We test our approach on the luminance components of several video sequences. The same character string containing

TABLE II
ANNOTATED EXCERPT OF CONTROL INFORMATION EXTRACTED FROM 660-FRAME WATERMARKED VIDEO SEQUENCE COMPRESSED AT 4.5 MBPS

Frame #	Video Content	Extracted Control Information			
		Rate Type for User Data Zero/Low/High	Frame Synch Index	# of bits @ high robustness	# of bits @ high payload
0	Flower f0	High	0	24	64
1	f1	High	0	24	64
2	f2	High	0	24	64
3	f3	High	0	24	64
4	f4	undecided	0	n/a	n/a
5	f5	High	0	24	64
6	f6	High	1	24	64
7	f7	High	1	24	64
...	...	...	...	...	...
142	f142	High	7	50	64
143	f143	High	7	50	64
144	f144	High	0	40	64
145	f145	High	0	40	64
146	f146	High	0	40	64
147	f147	High	0	40	64
148	f148	High	0	40	64
149	f149	High	0	40	64
150	Football f0	Zero	-1	0	0
151	f1	Zero	-1	0	0
152	f2	Zero	-1	0	0
153	f3	Zero	-1	0	0
154	f4	Zero	-1	0	0
155	f5	Zero	-1	0	0
156	f6	Low	1	4	8
157	f7	Low	1	4	8
...	...	...	...	...	...
364	T.Tennis f4	High	1	18	32
365	f5	High	1	18	32
366	f6	High	2	12	32
367	f7	High	2	12	32
...	...	...	...	...	...
448	f88	Zero	-1	0	0
449	f89	Zero	-1	0	0
450	f90	Low	7	4	8
451	f91	Low	7	4	8
...	...	...	...	...	...
TOTAL					
660 frames, 3 concatenated seq.			110 segments	1266 bits	3032 bits

Low confidence when extracting rate type info. from this B-frame due to compression.

Update synch index & embed new sets of user data in this new segment: bit25-48 @ high robustness, and bit65-128 @ high payload level.

Synch index updated from 7 to 0 in an 8-stage round robin fashion.

Repeatedly embed the same user payload in each frame of a segment (same synch).

No user data are embedded for a rather smooth segment. Nor is frame synch index embedded (as denoted by –1).

A small, predetermined amount of user data are embedded in segments with moderate achievable payload to reduce overhead.

Different segments of the same video sequence have different embedding capabilities.

access control information without error correction coding is hidden in two embedding levels. Between scene changes, we use equal-length segments, each containing six consecutive frames. One test video is the first 60 frames of the "flower garden" sequence, which has a frame size of 352×240 and a frame rate of 30 frames per second. The average PSNR of the watermarked video with respect to the original host signal is 32.5 dB. After data hiding, the video is encoded using MPEG-2, with a GOP structure of *IBBPBBI*. 18 characters (132 bits) can be extracted accurately when the video is compressed to 1.5Mbps or higher bit rate. An additional, longer string of 91 characters (640 bits) can be successfully extracted when compressed to 4.5Mbps or higher. Fig. 8 shows the 1st and 30th frames of the original and watermarked frames as well as their difference amplified by a factor of 5.

We also tested a longer and more diverse sequence of 660 frames by concatenating the "flower garden," "football," and "table tennis" sequences. A total of 3032 bits are embedded at high payload level and 1266 bits at high robustness level. All 4298 bits can be extracted accurately after 4.5 Mbps MPEG-2 compression or higher. When the video is compressed at 1.5 Mbps, the 1266 bits at high robustness level can still be correctly extracted, though the detector shows low detection confidence on 3 bits (0.2%). Error correction coding can be incorporated to correct a small percentage of errors. In Table II, an annotated excerpt of detection log shows the extracted control information and demonstrates the role of these control data for data embedding in diverse video sequences. We see that 1) repeatedly embedding the same payload in a few consecutive frames, together with frame sync index, help to combat occasional detection errors in severely distorted frames and 2) the adaptive embedding rate and the associated variable-length-encoded control information are effective in handling the uneven embedding capabilities across video segments. In addition to the user data payload and the associated control data, a constant spread-spectrum watermark signal sharing approximately 1/4

TABLE III
SUMMARY OF EXPERIMENTAL RESULTS FOR THE PROPOSED MULTI-LEVEL IMAGE AND VIDEO DATA HIDING SYSTEMS

	Level-1: high payload		Level-2: high robustness		Notes
	embedding rate	robustness	embedding rate	robustness	
512x512 lenna	32x32 pattern (1024 bits)	JPEG Q ≥ 45%, moderate additive noise.	"PINTL" (35 bits)	JPEG Q ≥ 20%; low pass filtering; additive noise.	PSNR = 42.5 dB
512x512 baboon			"Panasonic Tech." (105 bits)		PSNR = 33.6 dB
60-frame 352x240 flower garden sequence	640 bits (91 char.)	Mpeg-2 4.5Mbps; frame dropping	132 bits (18 char.)	Mpeg-2 1.5Mbps; frame dropping	Also hide control bits to facilitate extracting user data. Average PSNR is 32.5dB for flower garden.
660-frame 352x240 concatenated video sequence	3032 bits		1266 bits		

of JND's energy is embedded in every frame, which can be used to indicate ownership information and/or served as reference for image registration when the video encounters geometric distortion. The experimental results of both image and video data hiding are summarized in Table III.

IV. CONCLUSIONS

In this Part II, we demonstrate how the general solutions to the fundamental issues of data hiding presented in Part I can be used for specific design problems and applications. We have made extensive use of the two major types of embedding, the modulation and multiplexing techniques for embedding multiple bits, as well as shuffling for handling uneven embedding capacity. Using multilevel data hiding in image and video, we have shown that the amount of extractable information can be adapted to the actual noise conditions, making it attractive for unequal error protection on the embedded data and for progressive and scalable embedding.

REFERENCES

[1] M. Wu and B. Liu, "Data hiding in image and video: Part I—Fundamental issues and solutions," *IEEE Trans. Image Processing*, pp. –, June 2003.

[2] M. Wu, H. Yu, and A. Gelman, "Multi-level data hiding for digital image and video," in *Proc. Photonics East Conference on Multimedia Systems and Applications*, vol. 3845, Boston, MA, 1999.

[3] M. Wu and H. Yu, "Video access control via multi-level data hiding," in *IEEE Int. Conf. Multimedia & Expo (ICME'00)*, New York, 2000.

[4] H. A. Peterson, A. J. Ahumada, and A. B. Watson, "Improved detection model for DCT coefficient quantization," *Proc. SPIE*, vol. 1913, pp. 191–201, Feb. 1993.

[5] A. B. Watson, "DCT quantization matrices visually optimized for individual images," *Proc. SPIE*, vol. 1913, pp. 202–216, Feb. 1993.

[6] M. Wu and B. Liu, "Watermarking for image authentication," *Proc. IEEE Int. Conf. Image Processing (ICIP'98)*, 1998.

[7] F. Mintzer and G. Braudaway, "If one watermark is good, are more better?," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 4, Phoenix, AZ, Mar. 1999.

[8] C.-S. Lu and H.-Y. M. Liao, "Multipurpose Watermarking for Image Authentication and Protection," Tech. Rep., Institute of Information Science, Academia Sinica, R.O.C., 2000.

[9] A. Herrigel, J. Oruanaidh, H. Petersen, S. Pereira, and T. Pun, "Secure copyright protection techniques for digital images," in *Proc. 2nd Information Hiding Workshop (IHW)*, vol. 1525, Lecture Notes in Computer Science, 1998.

[10] H. V. Poor, *Introduction to Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.

[11] C. Podilchuk and W. Zeng, "Image adaptive watermarking using visual models," *IEEE J. Select. Areas Commun. (JSAC)*, vol. 16, May 1998.

[12] W. Zeng and B. Liu, "A statistical watermark detection technique without using original images for resolving rightful ownerships of digital images," *IEEE Trans. Image Processing*, vol. 8, pp. 1534–1548, Nov. 1999.

[13] D. Kundur, "Multiresolution digital watermarking: Algorithms and implications for multimedia signals," Ph.D dissertation, Univ. Toronto, Toronto, ON, Canada, 1999.

[14] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consumer Electron.*, vol. 38, no. 1, pp. 18–34, 1992.

[15] M. Wu, "Multimedia data hiding," Ph.D. dissertation, Princeton Univ., Princeton, NJ, 2001.

[16] H. Stone, "Analysis of attacks on image watermarks with randomized coefficients," NEC Research Institute, Tech. Rep., 96–045, 1996.

[17] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Digital Video Processing and Communications*. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[18] S. Pereira and T. Pun, "Fast robust template matching for affine resistant image watermarks," in *Proc. 3rd Information Hiding Workshop (IHW)*, Lecture Notes in Computer Science, 1999, pp. 207–218.

[19] G. Csurka, F. Deguillaume, J. J. K. ÓRuanaidh, and T. Pun, "A Bayesian approach to affine transformation resistant image and video watermarking," in *Proc. 3rd Information Hiding Workshop (IHW)*, Lecture Notes in Computer Science, 1999, pp. 315–330.

Min Wu (S'95–M'01) received the B.E. degree in electrical engineering and the B.A. degree in economics from Tsinghua University, Beijing, China, in 1996 (both with the highest honors), and the M.A. degree and Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1998 and 2001, respectively.

She was with NEC Research Institute and Signafy, Inc., in 1998, and with Panasonic Information and Networking Laboratories in 1999. Since 2001, she has been an Assistant Professor of the Department of Electrical and Computer Engineering, the Institute of Advanced Computer Studies, and the Institute of Systems Research at the University of Maryland, College Park. Her research interests include information security, multimedia signal processing, and multimedia communications.

Dr. Wu received a CAREER award from the U.S. National Science Foundation in 2002 and holds three U.S. patents on multimedia data hiding. She is a member of the IEEE Technical Committee on Multimedia Signal Processing and Publicity Chair of 2003 IEEE International Conference on Multimedia and Expo (ICME'03, Baltimore, MD). She co-authored a book *Multimedia Data Hiding* (New York: Springer-Verlag, 2002), and is a Guest Editor of special issue on Multimedia Security and Rights Management of the *Journal on Applied Signal Processing*.

Heather Yu (S'97–A'98) received the B.S. degree from Peking University, China, and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ, all in electrical engineering

She is a Senior Scientist at Panasonic Information and Networking Technologies Laboratory, Princeton. In 1998, she joined Panasonic and has since worked in the Security and E-commerce Group.

Dr. Yu is currently serving as Vice Chair of IEEE Communication Society Multimedia Technical Committee, Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA, Technical Program Chair for IEEE International Conference on Information Technologies, Research & Education (ITRE) 2003, Technical Program Chair for IEEE ICC 2004 Multimedia Symposium, Technical Program Vice Chair for IEEE International Conference on Multimedia and Expo (ICME) 2004. She served as Conference Program Chair, keynote speaker, panelist, panel chair, associate chair, session chair, technical committee member, steering committee member, and journal paper reviewer for many conferences and journals, in the field of multimedia processing and communication and information security. Her main research interests are in the areas of multimedia processing, rich media communication, and their applications in multimedia information access and distribution.

Bede Liu (S'55–M'62–F'72) was born in China and studied at the National Taiwan University (B.S.E.E., 1954) and the Polytechnic Institute of Brooklyn (D.E.E., 1960).

Prior to joining Princeton University, Princeton, NJ, in 1962, he had been with Bell Laboratories, Allen B. DuMont Laboratory, and Western Electric Company. He has also been a visiting faculty member at several universities in U.S. and abroad. He is Professor of electrical engineering at Princeton University. His current research interest lies mostly with multimedia technology, with particular emphasis on digital watermarking and video processing.

Dr. Liu and his students have twice received the Best Paper Awards on Video Technology (1994 and 1996). His other IEEE awards include Centennial Medal (1984) and Millennium Medal (2000), Signal Processing Society's Technical Achievement Award (1985) and Society Award (2000), Circuit and Systems Society's Education Award (1988) and Mac Van Valkenburg Award (1997). He is a member of the National Academy of Engineering. He was the President of the Circuit and Systems Society (1982), and the IEEE Division I Director (1984, 1985). He also served as the 1978 ISCAS Technical Program Chair and the General Chair of the 1995 ICIP.