Multiresolution Watermarking for Images and Video

Wenwu Zhu, Zixiang Xiong, and Ya-Qin Zhang

***Abstract*— This paper proposes a unified approach to digital watermarking of images and video based on the two- and three-dimensional discrete wavelet transforms. The hierarchical nature of the wavelet representation allows multiresolutional detection of the digital watermark, which is a Gaussian distributed random vector added to all the high-pass bands in the wavelet domain. We show that when subjected to distortion from compression or image halftoning, the corresponding watermark can still be correctly identified at each resolution (excluding the lowest one) in the wavelet domain. Computational savings from such a multiresolution watermarking framework is obvious, especially for the video case.**

***Index Terms*— Copyright protection, multimedia, watermarking, wavelet transforms.**

I. Introduction

With the rapid growth of network distributions of images and video, there is an urgent need for copyright protection against pirating. Different digital watermarking schemes have been proposed to address this issue of ownership identification. Early work on digital watermarking focused on information hiding in the spatial domain. For example, Schyndel *et al.* proposed to insert a watermark by changing the least significant bit of some pixels in an image [1]. Bender *et al.* described a watermarking approach by modifying a statistical property of an image [2]. Recent efforts are mostly based on frequency-domain techniques for still images. In particular, Cox *et al.* described a method where the watermark is embedded in large discrete cosine transform (DCT) coefficients using ideas borrowed from spread spectrum in communications [3], [4]. For digital watermarking of video sequences, Hartung and Girod [6] proposed a watermarking technique for MPEG-2 encoded video in the bitstream domain. Swanson *et al.* also considered MPEG-2 compressed domain video watermarking [7] and a wavelet-based multiresolution video watermarking method [8], in which the multiresolutional wavelet transform is performed in the temporal domain only. Although different transforms (e.g., discrete Fourier transform, discrete cosine transform, and discrete wavelet transform) have been used in digital watermarking schemes reported in the literature, there is no common framework for multiresolutional digital watermarking of both images and video.

In this paper, we propose a unified approach to digital watermarking of images and video based on the two-dimensional (2-D) and three-dimensional (3-D) discrete wavelet transforms [9]. Our wavelet-based watermarking framework is motivated by the fact that most network-based images and video are in compressed form and that wavelets are playing an important role in upcoming compression standards such as JPEG2000 and MPEG-4. We first experimentally show that a watermark signal [e.g., an independently identically distributed (i.i.d.) Gaussian random vector] can be embedded in *every* high-pass wavelet coefficient without any impact on the visual fidelity. This is different from the approach in [3], where the watermark is only placed into a small number of the perceptually most important coefficients (e.g., 1000 largest coefficients). Our results indicate that the capacity or the amount of information in an invisible watermark can be quite large.

We then describe the proposed framework where an i.i.d. Gaussian random vector is added to all the high-pass bands in the wavelet domain as a multiresolutional digital watermark. The watermark added to a lower resolution can be thought of as a nested version of the one corresponding to a higher resolution. The hierarchical nature of the wavelet representation allows detection of watermarks at all resolutions except the lowest one. Detection of lower resolution watermarks reduces computational complexity, as fewer frequency bands are involved. This computational savings can be quite significant for the video case.

The multiresolutional property makes our watermarking scheme robust to image/video downsampling operation by a power of two in either space or time. We also test our proposed watermarking scheme against common distortions introduced by compression and image halftoning. We use state-of-the-art wavelet image and video coders [10], [11] for compression and error diffusion for halftoning [12]. Experiments show that for both cases, the corresponding watermark can still be correctly identified at each possible resolution in the wavelet domain.

II. Capacity Issues in Digital Image Watermarking

Digital watermarking is a process of hiding a watermark (or signature) signal in image or video media by making small changes in the media content. Properties of watermarks include unobstructiveness and robustness. The former indicates that a watermark should be perceptually invisible; the later means that the watermark should be difficult to remove or destroy before resulting in severe degradation in visual fidelity. To make the watermark invisible, one would intuitively pick a watermark signal with small energy and hide it in the perceptually insignificant regions. However, the main thrust of [4] is the placement of the watermark in the perceptually significant regions of an image for robustness. It is argued that visual fidelity is only preserved if the perceptually significant regions

Manuscript received March 25, 1998; revised December 30, 1998. This paper was recommended by Associate Editor C. W. Chen.

W. Zhu is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA.

Z. Xiong is with the Department of Electrical Engineering, University of Hawaii, Honolulu, HI 96822 USA.

Y.-Q. Zhang is with Microsoft Research, Beijing 100080 China.

Publisher Item Identifier S 1051-8215(99)02961-4.

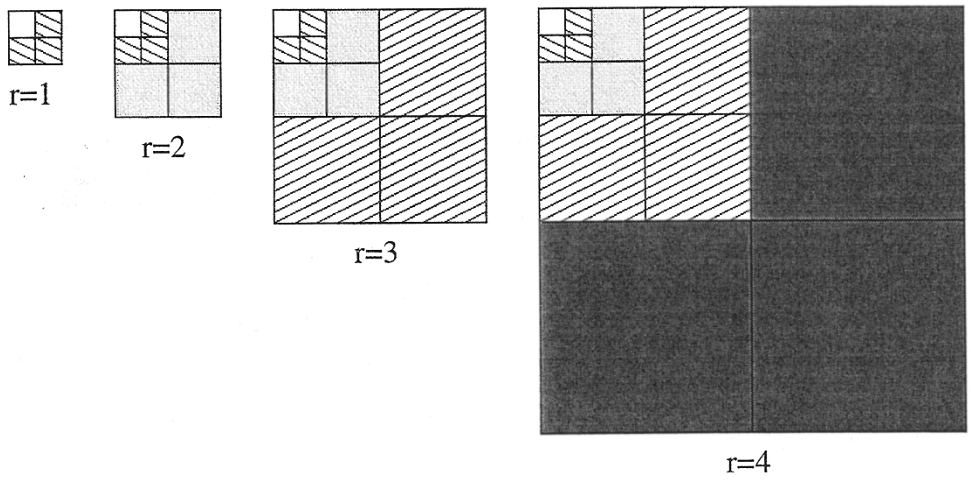

Fig. 1. Wavelet decomposition gives a multiresolutional representation of images. A digital watermark is added to the high-pass bands (shaded regions) at each resolution, and watermarks from low resolution to high resolution are nested.

Fig. 2. (a) The original 512 × 512 Lena image and (b) the watermarked Lena image. The watermark is a Gaussian distributed random vector added to all the high-pass bands in the wavelet domain. (c) The watermarked Lena image after SPIHT [10] compression (32 : 1). (d) The watermarked Lena image after Floyd–Steinberg error diffusion [12].

remain intact. Otherwise, a watermark placed in perceptually insignificant regions can be easily removed.

The question, therefore, is how much extra watermark information we can add to the perceptually significant regions without any impact on the visual fidelity. This raises the capacity issues in digital image watermarking. For a fixed watermarking procedure, the length of the watermark serves as a measure of the capacity, and it is upper bounded by the number of coefficients in the perceptually significant regions. A longer watermark signal means that more coefficients need to be modified; hence the watermarked images or video look noisier. The watermark will be perceptually visible beyond the capacity. For simplicity, in this paper we refer to capacity as *length capacity*.

TABLE I
PSNR (dB) BETWEEN THE ORIGINAL AND WATERMARKED IMAGES. WHEN THE WATERMARK LENGTH IS 1000, THE WATERMARK IS PLACED IN THE 1000 LARGEST HIGH-PASS WAVELET COEFFICIENTS; WHEN THE WATERMARK LENGTH IS 261 120, THE WATERMARK IS PLACED IN ALL HIGH-PASS WAVELET COEFFICIENTS

Watermark length	Lena	Barbara	Goldhill
1000	44.62 dB	44.24 dB	46.13 dB
261120	42.77 dB	40.53 dB	43.12 dB

To determine the perceptual capacity of each frequency, one could use models for the human visual system or simple experiments. Using the watermarking procedure described in Section III, we experimentally vary the length of the watermark signal and compute the peak signal-to-noise ratio (PSNR) between the original and watermarked images. Two cases are considered. First, the watermark is placed on the 1000 largest high-pass wavelet coefficients. Second, the watermark is hidden in all high-pass wavelet coefficients. The results are tabulated in Table I for three popular 512 × 512 images. We see that in all cases, the PSNR's of the watermarked images are quite high, making them visually indistinguishable from the originals even when a watermark (the longest) is placed in all high-pass wavelet coefficients.

It is beyond the scope of this paper to discuss the theoretical bound on watermark capacity, as it depends on the watermarking procedure, the source media, and the length of the watermark, among other things. However, our simple experiments seem to indicate that the watermark capacity is high enough to allow the longest watermark. High capacity facilitates robustness, as more watermark information can be hidden in the data. It also leads to the multiresolution watermarking framework described below.

III. A MULTIRESOLUTION WATERMARKING FRAMEWORK FOR IMAGES AND VIDEO

We describe our multiresolution watermarking scheme for images, assuming that a 2-D discrete wavelet transform is

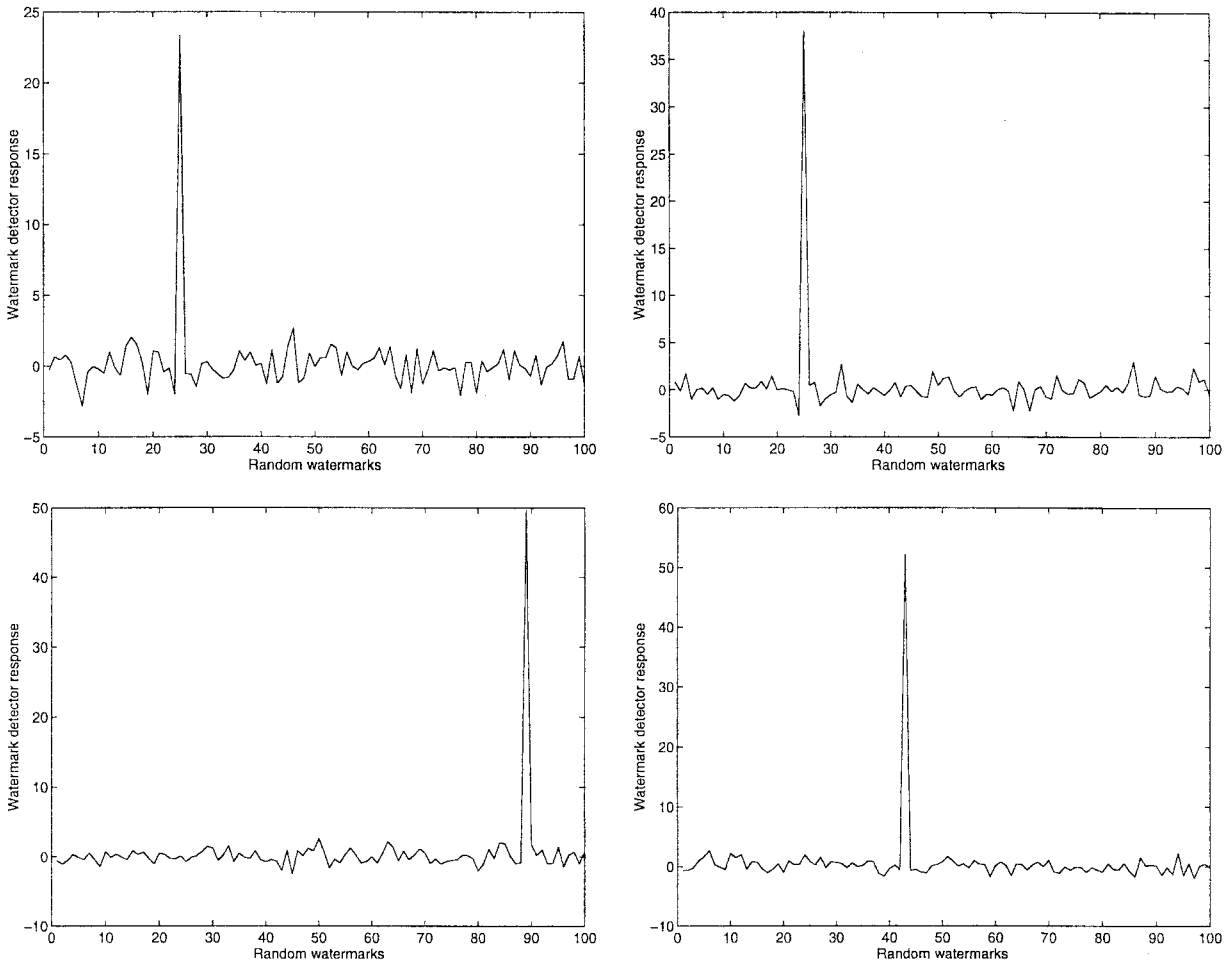

Fig. 3. Detector responses of watermarks X_r in four resolutions (size 64 × 64 to 512 × 512) of the watermarked Lena after compression and 100 randomly generated watermarks. In each resolution, only one peak exceeds the threshold $T\,(=6)$, which corresponds to the randomly generated watermark and is exactly the same as the original.

applied to an original image. Our proposed framework can be easily extended to video sequences after applying the 3-D discrete wavelet transform to a group of pictures (GOP). In the following derivation, we choose Cox's approach [3] because it is simple and well known. In fact, one may use many other algorithms as well (e.g., [13]).

Following the approach of [3], where the watermark is placed into the ac coefficients in the DCT domain, we add the watermark (an i.i.d. Gaussian random vector X) to all the high-pass bands in the wavelet domain. For the sake of simplicity, we assume that the image size is $N \times N$ and that there are in total $R+1$ resolutions in the wavelet image representation (this corresponds to an R-level wavelet transform). Let the watermark vector be

$$X = \{x_1, \cdots, x_{n_1}, \cdots, x_{n_2}, \cdots, x_{n_R}\}$$

where each element x in X is drawn independently according to $N(0, 1)$. Then the watermark corresponding to resolution r $(1 \leq r \leq R)$ is

$$X_r = \{x_1, \cdots, x_{n_r}\}$$

with $n_r = N^2/2^{2(R-r)} - N^2/2^{2R}$. The watermarks X_r at different resolutions are obviously nested, i.e., $X_1 \subset X_2 \subset \cdots \subset X_R$ (see Fig. 1).

To insert the watermark X into the high-pass wavelet coefficients, we collect them in a vector V and insert watermark X into V to obtain V' according to the formula [3]

$$v_i' = v_i(1 + \alpha_i x_i). \tag{1}$$

This nonlinear insertion procedure adapts the watermark to the energy present in each wavelet coefficient. It draws an analogy to spread spectrum in communications. The advantage of (1) is that when the coefficient v_i is small, the watermark energy is also small, thereby avoiding artifacts; and when v_i is large, the watermark energy is increased for robustness. The scaling constant α_i varies the energy/amplitude of the watermark and plays a role of balancing unobstructiveness and robustness. In our experiments, we fix α_i as a constant in all frequency bands for simplicity and choose $\alpha_i = 0.1$, as in [3].

To detect the watermark X' extracted from V', we first evaluate the detector response (or similarity of X and X')

$$\text{sim}(X, X') = X' \cdot X / \|X'\|_2$$

and then compare it with a threshold $T(= 6)$ to decide if the watermarks match. It was shown that $\text{sim}(X, X')$ follows

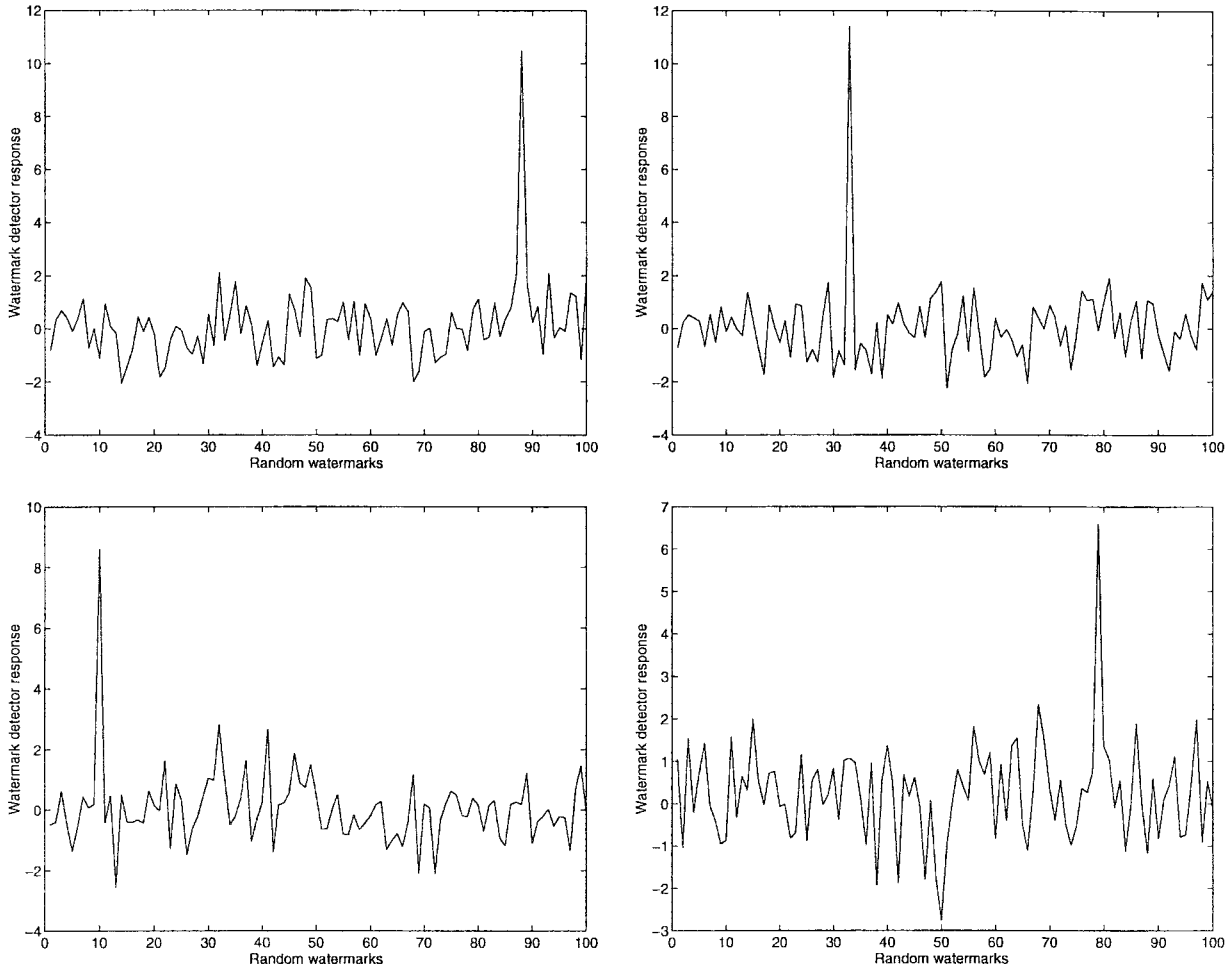

Fig. 4. Detector responses of watermarks X_r in four resolutions of the watermarked Lena after error diffusion and 100 randomly generated watermarks. In each resolution, again only one peak exceeds the threshold $T\,(=6)$.

the standard normal distribution with unit variance since X and X' are uncorrelated [3]. Setting the threshold $T = 6$ makes the probability of wrong watermark detection very small ($p = 1/\sqrt{2\pi}\int_6^\infty e^{-(1/2)t^2}\,dt = 9.8659 \times 10^{-10}$).

IV. Experiments

For still images (e.g., the 512 × 512 Lena image), we consider set partitioning in hierarchical trees (SPIHT) compression [10] and error diffusion [12] of the watermarked Lena image and test the effectiveness of our proposed multiresolution watermarking scheme. Fig. 2(a)–(c) shows the original Lena image, the watermarked Lena image with a PSNR of 42.77 dB, and the watermarked Lena image after SPIHT compression (32 : 1) with a PSNR of 33.48 dB, respectively. We choose a medium compression ratio of 32 : 1, as the visual quality for images is usually high in Web applications. Fig. 2(d) is the watermarked Lena image after Floyd–Steinberg error diffusion [12]. Detector responses of watermarks X_r in four resolutions (size 64 × 64 to 512 × 512) of the compressed Lena and 100 randomly generated watermarks are shown in Fig. 3. The single peak values of $\mathrm{sim}(X_r, X_r')$ at corresponding resolutions are 23.30, 37.97, 49.70, and 52.17. We can easily identify the watermark we added at each resolution, as these peak values far exceed the threshold $T\,(=6)$. These peak values indicate that the original watermark can be detected uniquely. We notice that the peak detector response increases with resolution. This is due to the fact that quantization noise in SPIHT is uniformly distributed over all frequency bands. When more frequency bands (or useful data) are involved in watermark detection at higher resolutions, it is easier to identify the correct watermark.

Fig. 4 shows the detector responses of watermarks X_r in four resolutions of the error-diffused Lena and 100 randomly generated watermarks. The single peak responses at four resolutions are 10.49, 11.42, 8.60, and 6.59, which are greater than $T\,(=6)$, ensuring the correct watermark identification at each resolution. We observe that in the case of error-diffusion, the peak detector response does not always increase with resolution. Instead, it decreases after a certain resolution. This is because of the blue noise [12] introduced in error diffusion, whose power increases with frequency. The three highest frequency bands of an error-diffused image will be dominated by the blue noise. This suggests that for halftone images, more weight should be put on detector responses at relatively low resolutions.

For video sequences, we test our proposed watermarking scheme against compression. After a watermark is inserted into

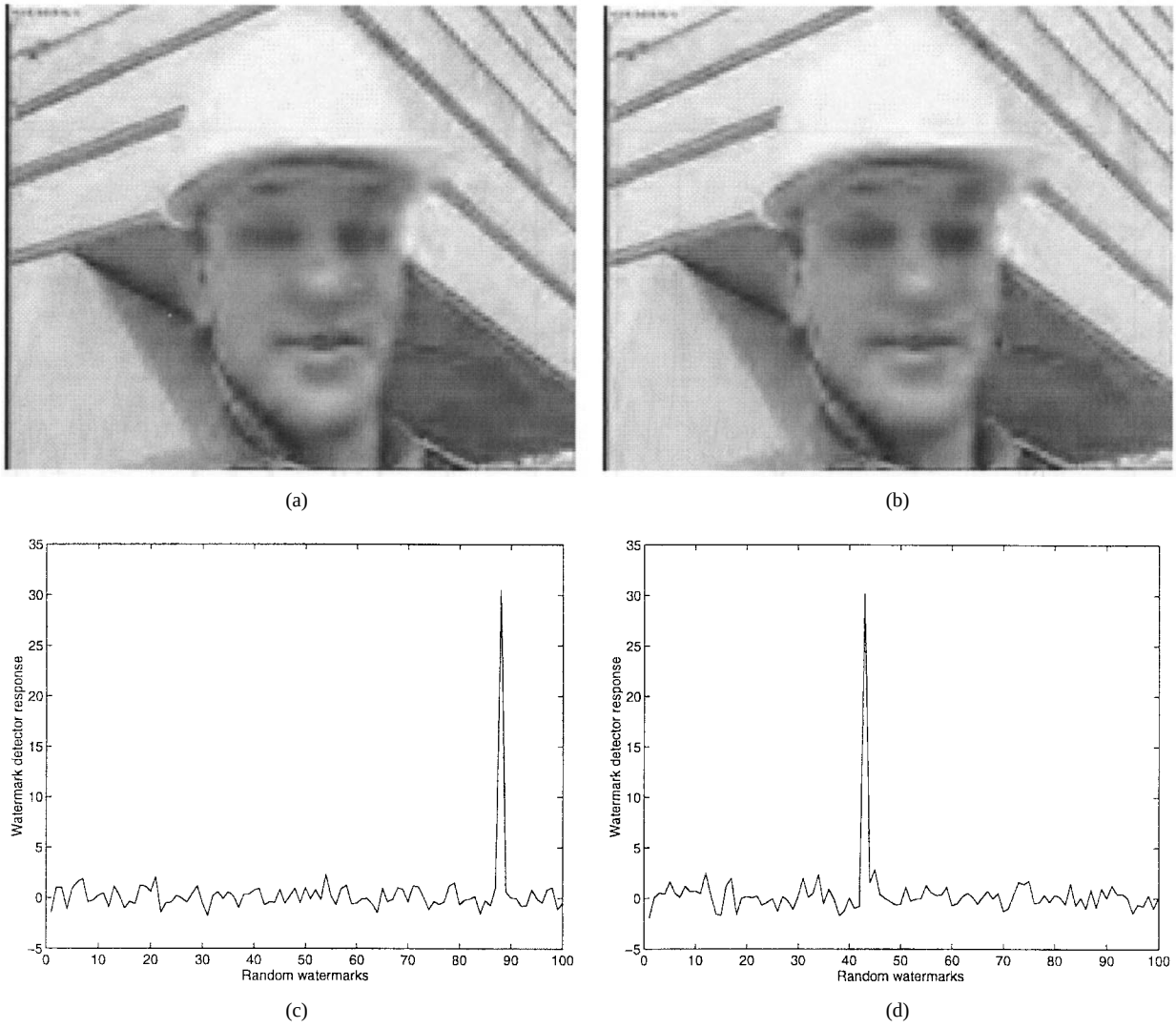

Fig. 5. Multiresolutional watermark detection after video compression. (a) First frame of the decoded QCIF *Foreman* sequence from the 3-D embedded SPIHT video coder in [11]. Encoding bit rate = 28.5 kb/s, frame rate = 10 f/s. (b) First frame of the watermarked QCIF *Foreman* sequence after 3-D SPIHT decoding. (c) Detector responses of watermark in full-resolution GOP ($176 \times 144 \times 16$) of the decoded *Foreman* sequence and 100 randomly generated watermarks. (d) Detector responses of watermark in half-resolution GOP ($88 \times 72 \times 8$) of the decoded *Foreman* sequence and 100 randomly generated watermarks. Again, the detector responses have only one peak greater than six in both resolutions.

a GOP (16 frames) of the original QCIF *Foreman* sequence, we use the 3-D wavelet-based embedded SPIHT video coder [11] to compress it at a bit rate of 28.5 kb/s and a frame rate of 10 f/s. A three-level wavelet transform is used in 3-D SPIHT, and the length of the watermark is 404 712 ($176 \times 144 \times 16 - 22 \times 18 \times 2$). The first frame of the decoded QCIF *Foreman* sequence is shown in Fig. 5(a), while the first frame of the watermarked QCIF *Foreman* sequence after 3-D SPIHT compression is shown in Fig. 5(b). Compared to the first frame of the original sequence, these two frames have PSNR's of 29.99 dB and 29.46 dB, respectively. Detector responses to 100 randomly generated watermarks are shown in Fig. 5(c) for full ($176 \times 144 \times 16$) resolution and in Fig. 5(d) for half ($88 \times 72 \times 8$) resolution. The single peak response in each resolution occurs when the randomly generated watermark is exactly the same as the original one. The peak value is 30.51 for full-resolution detection and 30.18 for half-resolution detection, guaranteeing correct watermark identification at both resolutions. Since half-resolution detection involves only one-eighth of the wavelet coefficients in the full-resolution case, computational savings is about 87.5%.

V. Concluding Remarks

We described a unified approach to digital watermarking of images and video using 2-D and 3-D wavelet transforms. Adding the watermark to all high-pass wavelet coefficients allows detection at different resolutions (excluding the lowest one) in the wavelet domain. Robustness of our watermarking scheme is tested against image/video compression and digital halftoning.

Our proposed approach can be easily extended to object-based watermarking by applying critically sampled shape-adaptive wavelet transforms [14, [15] where segmentation maps can be used to separate watermarks corresponding to different objects.

REFERENCES

[1] R. G. Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermarking," in *Proc. ICIP94*, 1994, vol. 2, pp. 86–90.

[2] W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," in *Proc. SPIE*, vol. 2420, 1995.

[3] I. Cox and M. L. Miller, "A review of watermarking and the importance of perceptual modeling," in *Proc. SPIE*, 1997, vol. 3016, pp. 92–99.

[4] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[5] X.-G. Xia, C. G. Boncelet, and G. R. Arce, "A multiresolution watermark for digital images," in *Proc. ICIP97*, Santa Barbara, CA, 1997, vol. I, pp. 548–551.

[6] F. Hartung and B. Girod, "Digital watermarking of MPEG-2 coded video in the bitstream domain," in *Proc. ICASSP97*, 1997, vol. 4, pp. 2621–2624.

[7] M. Swanson, B. Zhu, B. Chau, and A. Tewfik, "Object-based transparent video watermarks," in *Proc. IEEE Workshop Multimedia Signal Processing*, Princeton, NJ, June 1997, pp. 369–374.

[8] ———, "Multiresolution video watermarking using perceptual models and scene segmentation," in *Proc. ICIP97*, Santa Barbara, CA, 1997, vol. II, pp. 1151–1154.

[9] W. Zhu, Z. Xiong, and Y.-Q. Zhang, "Multiresolution watermarking for images and video: A unified approach," in *Proc. ICIP98*, Chicago, IL, Oct. 1998.

[10] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, June 1996.

[11] W. A. Pearlman, B.-J. Kim, and Z. Xiong, "Embedded video subband coding with 3D SPIHT," in *Wavelet Image and Video Compression*, P. Topiwala, Ed. Norwell, MA: Kluwer, 1998.

[12] R. A. Ulichney, *Digital Halftoning*. Cambridge, MA: MIT Press, 1987.

[13] I. Pitas, "A method for watermark casting on digital images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 775–780, Oct. 1998.

[14] J. Li and S. Lei, "Arbitrary shape wavelet transform with phase alignment," in *Proc. ICIP'98*, Chicago, IL, Oct. 1998.

[15] S. Li, W. Li, H. Sun, and Z. Wu, "Shape adaptive wavelet coding," in *Proc. IEEE ISCAS'98*, 1998, pp. V-281–V-284.