Watermarking of MPEG-4 Video Objects

Mauro Barni, *Member, IEEE*, Franco Bartolini, *Member, IEEE*, and Nicola Checcacci

***Abstract*—The recent finalization of MPEG-4 will make this standard very attractive for a large range of applications such as video editing, Internet video distribution, wireless video communications. Some of these applications are likely to get great benefit from watermarking technology, since it can enable a number of innovative services, such as conditional access policies, data annotation, data labeling, content authentication, to be implemented at a low price. One of the key points of the MPEG-4 standard is the possibility to access and manipulate objects within a video sequence. Thus object watermarking has to be achieved in such a way that, while a video object is transferred from a sequence to another, it is still possible to correctly access the data embedded within the object itself. The algorithm proposed in this paper embeds a watermark in each video object by imposing a particular relationship between some predefined pairs of quantized discrete cosine transform (DCT) coefficients in the luminance blocks of pseudo-randomly selected macroblocks (MBs). Watermarks are equally embedded into intra and inter MBs. Experimental results are presented validating the effectiveness of the proposed approach.**

***Index Terms*—MPEG-4 watermarking, objects watermarking, video watermarking.**

I. Introduction

DUE to the large diffusion of powerful personal computers and of wide band telecommunication networks the problem of illegal copying and distribution of digital contents has become very important in the last decade. On the other side, the advancement of these technologies is also seen as a big opportunity for dramatically increasing the dimension of the market of digital content, and to offer new services unimaginable until a few years ago.

The above aspects, and many others, are carefully taken into account by the ISO MPEG-21 initiative [1] whose goal is to achieve "an environment that is capable of supporting the delivery and use of all content types by different categories of users in multiple application domains." Among the technologies that MPEG-21 is investigating is that of persistent identifiers [2], i.e., of identification codes that are tightly and indissolubly attached to the content itself: these identifiers could be used for implementing a large number of services, ranging from the simple providing of detailed information regarding the content, to the protection of the content IPR, up to the implementation of conditional access policies. Data hiding, more precisely watermarking, can have an important role in providing the technological solution for the realization of such persistent identification. At the same time the recent finalization of MPEG-4 (a good overview can be found in [3]) makes this standard the natural tool to reach the goals indicated by the MPEG-21 standard, thus requiring that proper watermarking techniques are developed to fit the object structure of MPEG-4.

The principle of watermarking is to embed a digital code (watermark) within the host multimedia document, and to use such a code to prove ownership, to prevent illegal copying, or simply to give some indications about the watermarked data or to enable the access to enhanced versions of the content or to additional services. The watermark code is embedded by making imperceptible modification to the digital data. The embedded watermark must be resistant to the processing a video sequence is commonly submitted to. One of the key points of MPEG-4 video coding is the possibility to access and manipulate objects within a video sequence directly in the compressed domain. Thus object watermarking has to be achieved in such a way that, while a video object is transferred from a sequence to another (object manipulation), it is still possible to correctly access the watermark contained in the object itself.

Essentially, the structure of MPEG-4 coding is not different from previous video standards such as MPEG-1 and MPEG-2, in that block-based motion compensation and motion-compensated hybrid DPCM/transform coding techniques are used, the main difference is that coding is content-based, i.e., single objects are coded individually. Each frame of an input sequence is segmented into a number of arbitrarily shaped regions, Video Object Planes (VOPs), and the shape, motion and texture information of the VOPs belonging to the same Video Object (VO) are coded into a separate Video Object Layer (VOL). The first VOP of a Group Of Video object planes (GOV) is coded intraframe (I-VOP coding mode) by splitting it into macroblocks (MB). Each MB contains luminance and chrominance 8×8 pixel blocks $f(x,y)$ (e.g., in the 4:2:0 format four luminance and two chrominance 8×8 blocks), these are processed through discrete cosine transform (DCT), producing the DCT blocks $F(u,v)$, and then quantized resulting in the $QF(u,v)$ blocks. Finally, an efficient prediction of the dc- and ac-coefficients of the DCT is performed and the corresponding prediction error blocks $PQF(u,v)$ computed. Prediction error blocks are zig-zag scanned and entropy coded to produce the bit-stream. Each subsequent VOP in the GOV is coded using interframe VOP prediction (P-VOP or B-VOP), i.e., it is motion compensated, and the residual prediction error signal is split into MBs, and then into blocks which are compressed in the same way as I-VOP blocks.

Manuscript received August 30, 2001; revised April 18, 2003. This work was supported in part by Grants from the Italian Ministry of School, University and Research (MIUR). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sankar Basu.

M. Barni is with the Dipartimento di Ingegneria dell'Informazione, Università di Siena, 53100 Siena, Italy (e-mail: barni@dii.unisi.it).

F. Bartolini, deceased, was with the Dipartimento di Elettronica e Telecomunicazioni, Università di Firenze, Florence, Italy.

N. Checcacci was with the Dipartimento di Elettronica e Telecomunicazioni, Università di Firenze, Florence, Italy. He is now with Telecom Italia, Roma, Italy.

Digital Object Identifier 10.1109/TMM.2004.840594

1520-9210/$20.00 © 2005 IEEE

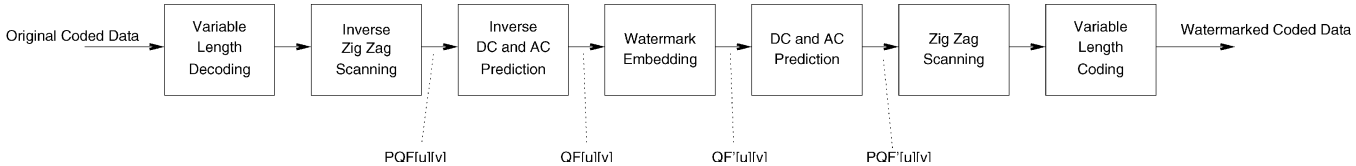


Fig. 1. Texture decoding process on a MPEG-4 decoder, modified to hide the watermark.

To be useful a digital video watermarking system must satisfy some basic requirements:

- **The embedded watermark should be perceptually invisible;** in other words its presence should not affect the video quality.
- **The watermark should be robust against common processing tools which do not seriously degrade the quality of the image;** for example digital video is usually stored in compressed format (like MPEG) by using lossy compression algorithms.
- **The embedded watermark should be robust against manipulations like cutting one or more frames of the video;** to achieve this goal it is necessary to insert the watermark information continuously in the video sequence (in other words every frame of the video should be watermarked).

The algorithm proposed in this paper embeds a watermark in each video object of an MPEG-4 coded video bit-stream by imposing specific relationships in a way similar to the system presented in [4] and [5]. In particular a relationship is imposed, if not naturally occurring, between some predefined pairs of quantized DCT middle frequency coefficients in the luminance blocks of pseudo-randomly selected MBs. Here, the adaptation of such a technique to the case of video objects watermarking is presented. An innovative approach to watermark recovery which exploits the data obtained from the whole sequence to improve reliability, and it is able to give a measure of the confidence of watermark reading, is also presented. The overall scheme of the proposed watermarking system is shown in Fig. 1. Quantized coefficients are recovered from the MPEG-4 bit-stream by reversing run-level entropy coding, zig-zag scanning and intra-dc and ac DCT coefficients prediction; DCT coefficients are modified to embed the watermark and then encoded again. In Fig. 1, we assume that the system is fed with an MPEG-4 video stream. Of course, if the system operates directly within the MPEG-4 coding chain, the first steps can be skipped.

The watermark is embedded both into INTRA and INTER MBs. A masking method is also adopted to limit visual artifacts in the watermarked VOPs and to improve, at the same time, the robustness of the system. Watermark recovery does not require the original video and is performed directly in the compressed domain.

The paper is organized as follows. In Section II, an overview of the main video watermarking algorithms developed until today is given; in Sections III and IV, the watermark embedding and retrieval algorithms are described respectively. In Section V, experimental results are presented aiming at demonstrating the validity of the proposed method. Finally, in Section VI, some conclusions are drawn.

II. Video Watermarking

The overwhelming majority of video watermarking algorithms developed so far, does not deal specifically with video objects watermarking. Thereby, in this section we will present the general state of the art of video watermarking, without focusing on object watermarking issues.

Existing video watermarking methods can be classified into two main classes according to the type of content the watermarking method works on: before (raw-video watermarking) or after compression (bit-stream watermarking).

Hartung and Girod's work for the watermarking of raw video [6] belongs to the first class of methods. According to their approach, inspired by spread-spectrum communications, a watermark, represented by a binary string, is lengthened so to fit the length of host data; the binary string is then modulated by a pseudo-noise sequence and then pixelwise added to the line-scanned luminance component of the video. Watermark recovery is accomplished by using a correlation-based method. The approach is, in general, sensitive to frame cutting and exchange. A variant in which the video sequence is modeled as a two-dimensional pixels plane (*bitplane*) is proposed in [7].

Hsu and Wu [8] address the raw-video watermarking issue by performing a DCT transform, on a block-by-block basis, of the video. In fact, to be resistant to MPEG temporal prediction, the watermark is inserted by imposing particular relationships between corresponding middle frequency DCT coefficients belonging to spatially neighboring blocks in intracoded frames or belonging to "*temporal*" neighboring blocks in intercoded frames (one block is chosen in the predicted frame and one in the reference frame). Finally, to obtain the watermarked video sequence, the inverse DCT of every frame is computed. In the proposed method, the GOP structure adopted by the MPEG coder is assumed to be known in advance. The extraction of the embedded watermark requires the watermarked video sequence and the original one.

Another interesting raw-video watermarking method has been proposed by Swanson *et al.* [9]. In their work, the authors present a method in which the watermark is inserted in the static and dynamic temporal components generated from a temporal wavelet transform of the video. The temporal and frequency distributions of the watermark are controlled by the masking characteristics of the host video signal, to obtain a reasonable tradeoff between visibility and robustness. The wavelet coefficient frames with the embedded watermark are converted back to the temporal domain using the inverse wavelet transform. In the original work by Swanson *et al.*, watermark recovery requires the original video sequence.

The raw-video watermarking system proposed by Kalker *et al.* [10] considers the video as a sequence of still images. The sequence is, then, marked by inserting the same watermark in all the frames. The embedding process is performed by adding samples of the watermark pattern, which are independently drawn from a normal distribution with mean and standard derivation equal to 0 and 1, respectively, to the pixel values of the considered frame. In other words the watermark is simply additive white noise. Basically, watermark detection is performed by spatial correlation.

Deguillaume *et al.* [11] proposed a three-dimensional (3-D) spatio–temporal DFT watermarking scheme in which the raw video is viewed as a 3-D signal with two dimensions in space and one dimension in time. The video is first divided into consecutive chunks of fixed length and then a 3-D DFT transform is performed on every chunk. Finally two kinds of information are hidden in the magnitude of every transformed chunk: a watermark and a template. The watermark, which is the same for every chunk, is a spread-spectrum signal representing the author signature which is added to the magnitude of the 3-D DFT. The template is a 3-D grid which is embedded into the 3-D DFT magnitude to determine and invert geometric transformations suffered by the video. Recovering of hidden information does not require the original video.

Hartung and Girod also proposed a bit-stream watermarking algorithm [12] that works in a way similar to the already mentioned frame watermarking method by the same authors. The watermark, consisting of a spread spectrum signal, is DCT transformed. The DCT coefficients of the watermark are then added to the nonzero DCT coefficients of the MPEG-2 coded video bit-stream by paying attention not to increase the bit rate. Watermark recovery requires the raw video (in other words in the case of MPEG-2 compressed videos the sequence must be decompressed before extracting the mark) and is performed by using a correlation-like method.

Another interesting bit-stream watermarking method has been presented by Langelaar *et al.* [13], [14], where the authors propose to mark only the intracoded frames (I-frames) of the MPEG video stream. Each bit of the watermark (the watermark is a binary string) is embedded in a region of 16 8×8 blocks by introducing an energy difference between the sets of high frequency DCT coefficients of the upper half and the lower half of the image region itself; the value of the embedded bit is defined by the sign of the energy difference introduced. This difference is obtained by discarding those DCT coefficients that in the zig-zag scan are located after a *cutoff* point and belong to one of the two mentioned half-regions. Watermark recovery is simply accomplished by evaluating the energy difference sign in the selected regions.

Video watermarking algorithms as those presented above do not consider the case of object watermarking, and, usually, they do not exploit the information derived from the whole sequence for extracting the embedded data. Furthermore, often they do not produce a measure of the confidence of reading, or, when such a confidence is available, it is not exploited to decide whether the video is watermarked or not. The aim of the work presented in this paper was just to overcome these limitations.

Actually, an MPEG-4 specific video watermarking algorithm has already been proposed by Piva *et al.* in [15]. A DWT-based watermarking algorithm originally developed for still images [16], [17] is applied to the watermarking of single MPEG-4 objects. The algorithm operates frame by frame by adding a pseudo-random watermark to the high-resolution bands of each object. Watermark concealment is improved by weighting the watermark by a proper masking function. Watermark recovery is based on the correlation between the watermark the detector is looking for and the DWT coefficients of the possibly marked objects. A drawback with the system proposed in [16] and [17] is that the detector can only reveal the presence of a known watermark (detectable watermarking or 1-bit watermarking [18]), thus limiting the watermark payload. The system proposed in this paper overcomes this problem, since it belongs to the class of readable watermarking algorithms [18], i.e., the decoder can effectively read the bits conveyed by the watermark without knowing them in advance (as it would have been necessary if the decoder could only decide whether the video sequence contains a given watermark or not).

III. Watermark Embedding

The watermarking algorithm proposed here hides a bit of the watermark code in every luminance block belonging to a set of MBs selected on a pseudo-random basis. As such the algorithm is only able to work when an MPEG4 bit-stream is available. If the MB is a skipped one the bit is also skipped. The watermarking code is repeated over the whole VOP (i.e., after the last bit of the code has been embedded, the process considers the first bit again, and so on). Watermark embedding is performed on a frame basis; that is, on every VOP of the same VOL the code is embedded again by starting from the first bit. Every bit is thus embedded more than once during a sequence of VOPs, but, due to MBs skipping, some bits are embedded less frequently than others.

In Fig. 2, the watermark embedding scheme is shown. The watermark is embedded in the video through the following steps:

1) select MBs and the DCT quantized coefficients pairs to be modified;
2) for each block belonging to a selected MB:
    a) compute the frequency mask;
    b) use the mask to weigh the watermark amplitude;
    c) modify, according to the algorithm rule, selected pairs of coefficients creating the watermarked block.

At the start of each VOP, a pseudo-random binary sequence is generated, based on a secret key and on the characteristics (number of MBs) of the VOP itself, for choosing those MBs where the watermarking code has to be embedded and the coefficients pairs to be modified. If the chosen MB is not skipped, one bit of the watermark code is embedded within it by imposing a particular relationship between the coefficients of selected pairs of coefficients that belong to each luminance block of the MB. If the MB is a skipped one, the bit is skipped too. The use of the pseudorandom sequence permits to improve the security of the watermark, by preventing the possibility for an attacker to alter the watermarking code after having identified the positions where the watermark was embedded. This makes the watermarking algorithm private [18], i.e., only allowed people

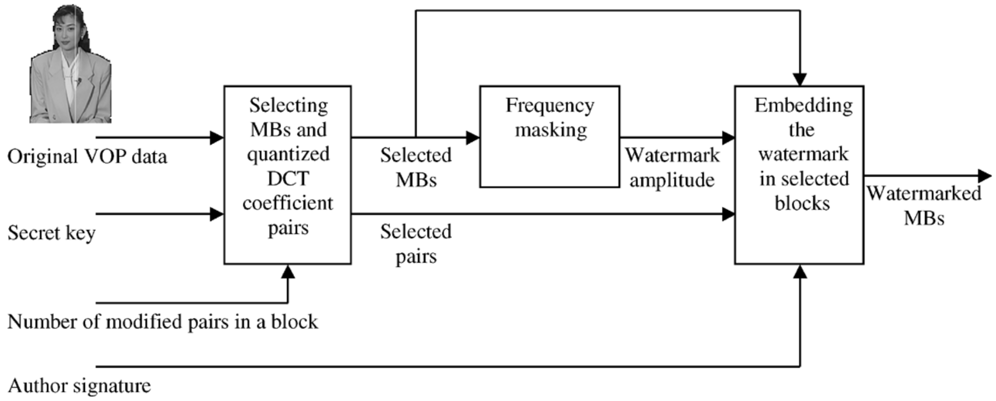
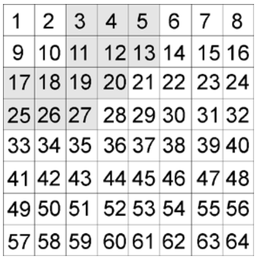

Original VOP data
Secret key
Number of modified pairs in a block
Author signature
Selecting MBs and quantized DCT coefficient pairs
Selected MBs
Selected pairs
Frequency masking
Watermark amplitude
Embedding the watermark in selected blocks
Watermarked MBs

Fig. 2. Diagram of the watermark embedding phase.



1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64

Fig. 3. Sketch of the mid-frequency quantized DCT coefficients that are watermarked.

(those who know the secret key) can read the watermark. Moreover, it is widely known that applying an identical watermark to each frame of a video leads to problems of maintaining statistical invisibility. To prevent statistical attacks (e.g., temporal averaging), pseudo random sequence generations it is also VOP dependent (i.e., the locations where the watermark is embedded change from VOP to VOP according to the dimension of the VOP itself); this is achieved by seeding the pseudo random generator with the sum of the secret key and the dimension in MBs of the VOP to be marked.

To achieve a tradeoff between the requirement of invisibility (changes in the low-frequency components of a video signal are more noticeable than those in the high-frequency components) and robustness (in compression process, like MPEG4, the video signal can be considered as low-pass filtered), only quantized DCT coefficients ($QF(u,v)$) belonging to a mid frequency range are considered for watermark embedding (as depicted in Fig. 3).

Similarly to what is done in [4] and [5] the watermark is carried by the difference among the magnitudes of some selected pairs of quantized DCT coefficients belonging to the mid-frequency region sketched in Fig. 3, i.e.:

$$W(u_1,v_1,u_2,v_2) = |QF(u_1,v_1)| - |QF(u_2,v_2)| \quad (1)$$

where (u_1,v_1) and (u_2,v_2) are the coordinates of the two coefficients of one of these pairs. It is expected that $W(u_1,v_1,u_2,v_2)$ is a nonstationary random process having zero mean, and a moderate variance if the coefficients composing each pair are sufficiently close each other. If $\{QF(u_1,v_1), QF(u_2,v_2)\}$ is a randomly selected pair, let the corresponding watermarked pair be denoted by $\{QF'(u_1,v_1), QF'(u_2,v_2)\}$. In particular embedding is performed in such a way that the difference in (1) is greater than 0 if a 1 informative bit has to be conveyed, and is lower than 0 if the informative bit is 0. By supposing the bit to be embedded is 1, three cases can hold:

- both coefficients of the pair are non zero and the difference of their magnitude is $\geq nA_F$;
- both coefficients of the pair are non zero and the difference of their magnitude is $< nA_F$;
- one or both coefficients of the pair are zero.

In the first case no modification of the coefficients is performed. In the second case, the watermark is inserted with maximum strength: the sign of the coefficients is not changed, while the respective magnitude becomes

$$|QF'(u_1,v_1)| = (|QF(u_1,v_1)| + |QF(u_2,v_2)| + \frac{nA_F}{2} \quad (2)$$

(where / is an integer division) and

$$|QF'(u_2,v_2)| = \begin{cases} |QF'(u_1,v_1)| - nA_F, & \text{if } |QF'(u_2,v_2)| > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where A_F is a masking parameter modulating the watermark energy to improve the invisibility and the robustness of the watermark embedded into each VOP. More specifically, A_F is set by relying on the model proposed in [19], where the authors describe a method to change the watermarking strength according to the smoothness and edginess characteristics of the blocks to be marked.

The other parameter appearing in (3), that is n, is used to take into account the fact that the quantization step (QP) of a given coefficient can change from block to block in order to keep the bit-rate as constant as possible: Given that we work on quantized levels the modification of coefficients quantized with a large quantization step can be much more visible than that of coefficients quantized with a small quantization step. Thus n is increased when the quantization step decreases. Moreover, to achieve a reasonable tradeoff between the requirements of invisibility and robustness, it is necessary that the larger is the number (C_{num}) of pairs that we want to mark in a block the smaller is the value of n. The adaptation rule of n was obtained experimentally resulting in the following formula:

$$n = \begin{cases} \frac{15-4C_{num}}{QP}, & \text{if } (15 - 4C_{num}) > QP \\ 1 & \text{otherwise.} \end{cases} \quad (4)$$

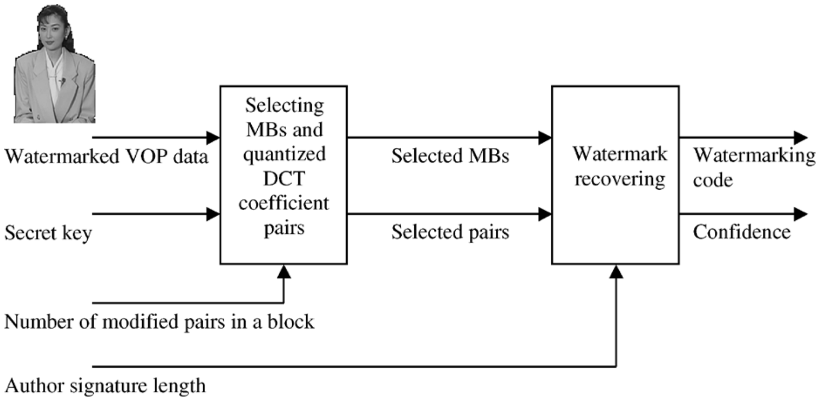


Fig. 4. Diagram of the watermark recovering phase.

When one or both coefficients of the DCT coefficients pair is zero, it is more difficult to maintain the watermark perceptually invisible because the masking effect between the DCT frequency components is absent. In this case, the coefficients of the pair are changed less heavily, in a way not to disturb the retrieval phase

$$QF'(u_1,v_1)=\begin{cases} QF(u_1,v_1), & \text{if } |QF(u_1,v_1)|>|QF(u_2,v_2)| \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$QF'(u_2,v_2)=\begin{cases} QF(u_2,v_2) & \text{if } |QF(u_1,v_1)|>|QF(u_2,v_2)| \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

For embedding a 0 bit, the algorithm is similar, but the roles of the coefficients (u_1, v_1) and (u_2, v_2) are exchanged, and thus, for the pairs where a 0 bit has been embedded, it results $W'(u_1, v_1, u_2, v_2) \leq 0$.

IV. Watermark Recovery

Watermark retrieval is carried out in two steps as it can be seen in Fig. 4.

The first step is analogous to the first step of the embedding process (see Section III) and requires the knowledge of the parameters used in the embedding phase (i.e., the secret key and the number of pairs that were modified in each selected block) to correctly identify MBs and coefficients pairs where the watermark was actually hidden. In the second step the relationships between the coefficients of the selected pairs are analyzed. The knowledge of the watermarking code length is needed to compute the repetition step of the watermarking code in each VOP (i.e., how many times the watermark was embedded in the considered VOP).

For reading the jth bit of the watermarking code, an accumulator Acc_j is considered, where the values of $W'(u_1, v_1, u_2, v_2)$ corresponding to all the pairs of coefficients where the bit was inserted, are summed. Let us call ψ_j the set of such pairs

$$Acc_j = \sum_{\psi_j} W'(u_1, v_1, u_2, v_2). \tag{7}$$

Such a sum is then compared to a threshold T_D, to evaluate the value of the embedded bit b_j:

$$b_j = \begin{cases} 1, & \text{if } Acc_j > T_D \\ \text{indeterminate}, & \text{if } -T_D \leq Acc_j \leq T_D \\ 0, & \text{if } Acc_j < T_D. \end{cases} \tag{8}$$

In order to minimize the overall error probability, the value of T_D should be set to 0 ; in fact, it is expected that Acc_j is positive when the embedded bit that is 1, and negative in the opposite case. On the other hand, choosing $T_D = 0$ leads to read a watermarking code also when no code was actually embedded. To obviate to this problem, a measure of the reliability (confidence) of bit reading is also provided for each recovered bit.

To this end we note that only one of the following situations is possible:

- $Hp.A$: the VOL is not marked;
- $Hp.B$: the VOL is marked.

Let us $E[Acc|A]$ and $\sigma^2_{Acc|A}$ denote, respectively, the mean value and variance of Acc conditioned to $Hp.A$. In the Appendix it is demonstrated that for each bit

$$E[Acc|A] = 0 \tag{9}$$

and an estimate of $\sigma^2_{Acc|A}$ based on the values of the DCT coefficients belonging to the nonwatermarked blocks (i.e., the blocks not selected by the random key) is given. Given $\sigma^2_{Acc|A}$, we define the confidence value for each bit as

$$C_j = \frac{Acc_j}{\sigma_{Acc|A}}. \tag{10}$$

The smallest is this value, the most probable is that the sequence is not watermarked. Small values are also obtained when the sequence is watermarked with another key: in this case the absolute value of the accumulator will be in general larger than for $Hp.A$, because some coefficient pairs where some bits are actually embedded can be selected, thus contributing to increase it. As a global measure of the possibility that a video is watermarked with the used key, the sum of the confidence values resulting for all the bits can be used. Note that the availability of a confidence measure for each bit, allows the use of soft decoding [20], [21] techniques for decreasing the BER (anyway the use of error correcting codes has not been contemplated in this paper).

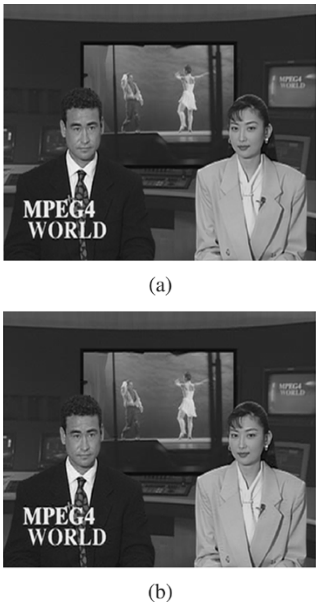


Fig. 5. Frames from “News” video sequence. (a) Original. (b) Watermarked.

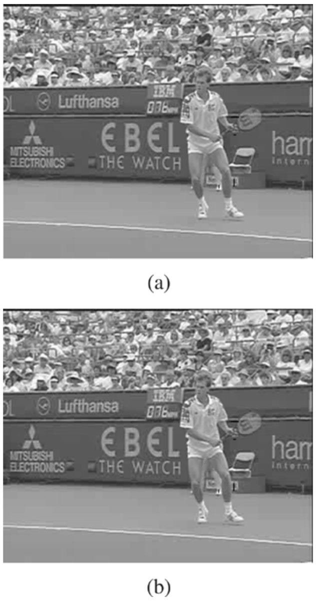


Fig. 6. Frames from “Stefan” video sequence. (a) Original. (b) Watermarked.

V. Experimental Results

In this section, some of the experiments carried out for proving the effectiveness of the proposed algorithm are discussed. The test software was implemented in such a way that it takes, in the embedding phase, an MPEG-4 VOL, a secret key (represented by an integer), the number of pairs that should be modified in a block, a binary watermarking code, and a floating point number representing the maximum strength of A_F, say A_{MAX}, of the watermarking signal, as input, it parses the VOL and writes it in a new file. In the recovery phase the implemented software takes a watermarked VOL, the watermarking code length, a secret key and the number of pairs that were modified in a block (that should be the same used in the embedding phase) as input, and gives the watermarking code and the confidence of each read bit.

Two kinds of tests were conducted on three different video sequences for proving on one side the invisibility of the embedded watermark and on the other side the robustness against all processing which does not seriously degrade the quality of the video. Among the tested video sequences the results regarding “News” and “Stefan” are presented here.

The standard video sequences “News” and “Stefan” were coded by using binary alpha planes and consists of 300 frames in CIF format with a frame rate of 25 fr/s. “News” is composed by four VOs [see Fig. 5(a)]: the anchor-man on the left is labeled as VO0, the woman on the right as VO1, the monitor in the center as VO2 and the background as VO3. The video sequence “Stefan,” instead, is composed by two VOs [see Fig. 6(a)]: the tennis player (VO1) and the background (VO0).

A. Visibility Experiments

To evaluate the quality of the watermarked videos a series of tests has been performed in which the original video and a video in which each VO is watermarked are displayed in sequence to a viewer. The order in which the original and the watermarked videos are displayed was randomly selected. The viewer was asked to select which of the sequences has better quality.

The embedded watermark appears perceptually undetectable and so each video was selected approximately 50% of the time. For the sake of completeness, two original frames of the two videos are shown in Figs. 5(a) and 6(a), even if the printing quality does not permit to fully appreciate the unobtrusiveness of the watermark. The corresponding watermarked frames are shown in Figs. 5(b) and 6(b).

B. Robustness Against Bit-Rate Decreasing

It is usually assumed that if each frame of a video is watermarked the removal of the watermark requires big computing power and memory resources [22]. Thus, in this paper, only manipulations which can be carried out by an average consumer (i.e., those requiring a limited computing power) are considered. The robustness against bit-rate decreasing is very important because this kind of attack can be either of intentional or “incidental” nature. In fact, in most application involving storage and transmission of digital video, it is necessary to reduce the bit rate in order to improve the coding efficiency. In the proposed test each watermarked video sequence is decoded and encoded again by decreasing the bit-rate, and watermark detection is attempted. This test is performed, in each video, for a set of different watermarking code lengths.

In Table I, the total number of correctly read bits is plotted against the bit-rate of the four VOs of the “News” sequence. It appears that in most cases free error decoding is achieved also when the bit-rate is halved with respect to the original (i.e., very low coding quality). A particular observation is worth for VO3 (the background), where only 15 bits can be reliably embedded due to the scarcity of motion in the scene, and thus to the low number of intercoded MBs available in the bit-stream for watermark embedding.

TABLE I
TABLES OF THE NUMBER OF CORRECTLY READ BITS FOR EACH VO OF THE "NEWS" SEQUENCE, AT DECREASING BIT RATE AND FOR THREE DIFFERENT VALUES OF THE CODE LENGTH (15, 20, AND 25 BITS)

VO0, VO1, VO2					
	Bit rate				
Code length	400	350	300	250	200
25	24	24	24	24	24
20	20	20	20	20	20
15	15	15	15	15	15

VO3				
	Bit rate			
Code length	300	250	200	150
25	22	21	20	19
20	19	19	18	17
15	15	15	15	15

TABLE II
TABLE OF THE NUMBER OF CORRECTLY READ BITS FOR VO0 AND V01 OF THE "STEFAN" SEQUENCE, AT DECREASING BIT RATE AND FOR FOUR DIFFERENT VALUES OF THE CODE LENGTH (15, 20, 25, AND 30 BITS)

VO0, VO1						
	Bit rate					
Code length	500	450	400	350	300	250
30	30	30	30	30	30	30
25	25	25	25	25	25	25
20	20	20	20	20	20	20
15	15	15	15	15	15	15

TABLE III
NUMBER OF CORRECTLY READ BITS FOR THE TWO VIDEO OBJECTS OF STEFAN WHEN 15 BITS ARE EMBEDDED, AND AFTER CUTTING THE FIRST FRAME OF THE VIDEO AND RE-ENCODING IT

	$A_{MAX} = 1.0$	$A_{MAX} = 3.0$
VO0 (1000 kbit/s)	15	15
VO1 (500 kbit/s)	15	15

TABLE IV
NUMBER OF CORRECTLY READ BITS FOR THE TWO VIDEO OBJECTS OF STEFAN WHEN 30 BITS ARE EMBEDDED, AND AFTER CUTTING THE FIRST FRAME OF THE VIDEO AND RE-ENCODING

	$A_{MAX} = 1.0$	$A_{MAX} = 3.0$
VO0 (1000 kbit/s)	28	30
VO1 (500 kbit/s)	30	30

TABLE V
NUMBER OF CORRECTLY READ BITS FOR THE FOUR VIDEO OBJECTS OF NEWS WHEN 15 BITS ARE EMBEDDED, AND AFTER CUTTING THE FIRST FRAME OF THE VIDEO AND RE-ENCODING

	$A_{MAX} = 1.0$	$A_{MAX} = 2.0$
VO0 (400 kbit/s)	15	15
VO1 (400 kbit/s)	14	15
VO2 (400 kbit/s)	15	15
VO3 (300 kbit/s)	13	14

TABLE VI
NUMBER OF CORRECTLY READ BITS FOR THE FOUR VIDEO OBJECTS OF NEWS WHEN 20 BITS ARE EMBEDDED, AND AFTER CUTTING THE FIRST FRAME OF THE VIDEO AND RE-ENCODING

	$A_{MAX} = 1.0$	$A_{MAX} = 2.0$
VO0 (400 kbit/s)	20	20
VO1 (400 kbit/s)	17	19
VO2 (400 kbit/s)	20	20
VO3 (300 kbit/s)	14	16

In Table II, the results for the "Stefan" sequence are presented. In this case up to 30 bits can be reliably embedded, thanks to the larger number of MBs available for embedding.

C. Robustness Against Frame Dropping

Frame drops may arise either intentionally or not. The encoding process may in fact result in frame skipping; similarly, in videos with very low motion components (i.e., high interframe correlation) frame cutting can be performed without significantly degrading the quality. The proposed watermarking algorithmX inserts the watermark without modifying zero coefficients; for this reason the strength of the embedded watermark is higher in INTRA coded VOPs than in INTER coded VOPs. Changes in the GOV structure (i.e frame cutting) are then critical; it is also obvious that one of the worst cases is obtained when the first frame of the video is dropped. In that case, every INTRA coded VOP becomes the last INTER coded VOP in the GOV structure (i.e., coarsely quantized and heavily affected by motion compensation errors).

In the proposed test each watermarked video sequence is decoded and encoded again (at the same bit-rate) after cutting the first frame, and watermark detection is attempted. In Tables III–VI, the number of correctly read bits for each video object is shown against A_{MAX} for two different watermarking code lengths. Although such a manipulations is a very dangerous one, at least 15 bits can still be reliably hidden within almost all the video objects. The most critical video object is again the background of "News."

D. Confidence Measure

Regarding the confidence measure defined by (10), in Fig. 7 the absolute values of the confidence estimates obtained for VO 0 of the "Stefan" sequence are plotted for each bit: the three cases of reading the watermark from a watermarked copy using the correct key (label "Watermarked"), reading the watermark from a non watermarked copy (label "Non Watermarked"), and reading the watermark from a watermarked copy by using the wrong key (label "Wrong Key") are considered. It is evident that the confidence values are always higher when the correct key is used on a watermarked sequence than in the other two cases. For some of the bits (e.g., bit 0 and bit 17) the value of the confidence is quite low also when the correct key is used: this is due to the fact that these particular bits have been repeated only a few times in the sequence, and thus their reading is quite unreliable. These are also the two bits that are lost when the first frame of the sequence is cut (see Table IV). Furthermore it is worth observing that the confidence estimate results to be usually higher for the "Wrong Key" case than for the "Non Watermarked" case: this can be explained by the fact that, from time to time, some watermarked pairs are selected to contribute to Acc, even if the key is wrong. Anyway the average of the confidence values of all the bits results to be 13.05 for the "Watermarked" case, 1.83 for the "Non Watermarked" case, and 3.41 for the "Wrong Key" case: it can, thus, be assumed to be a good parameter for deciding if the sequence is really watermarked with that key or not.

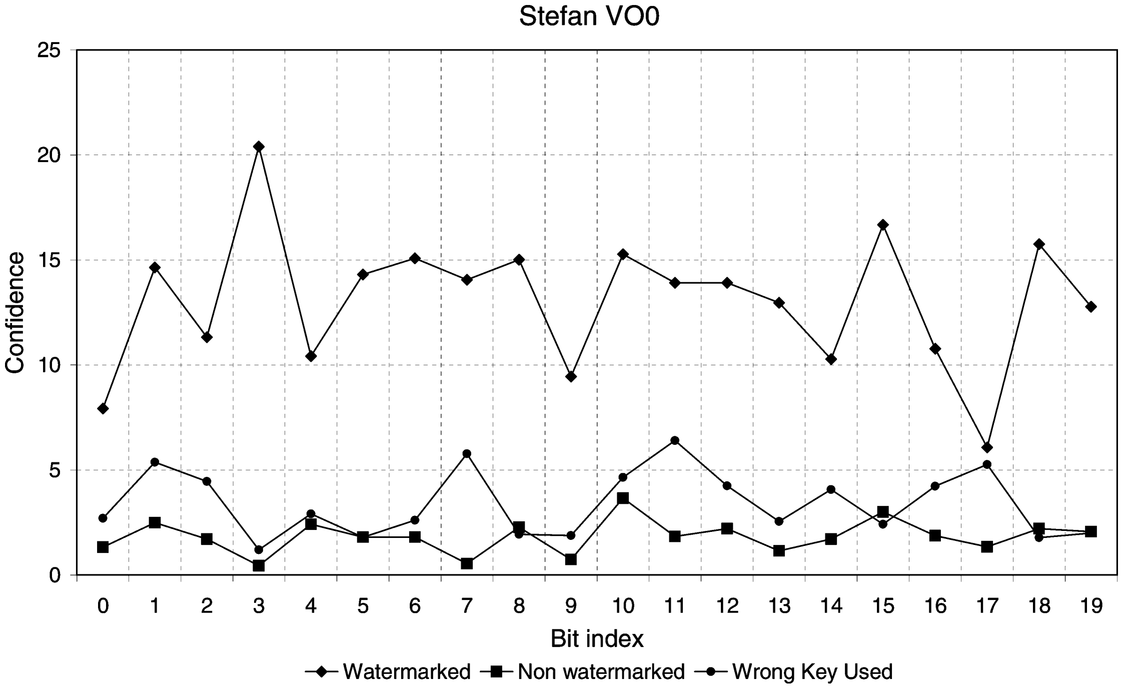


Fig. 7. Plot of the confidence measure obtained for each of the 20 bits embedded into the "Stefan" sequence.

VI. Conclusions

Driven by the growing interest toward the MPEG-4 video coding standard, we have presented a scheme for the watermarking of MPEG-4 video objects. The proposed algorithm works directly in the compressed domain thus reaching a high degree of flexibility and ease of use. The possibility of distinguishing between marked and nonmarked contents is also envisaged. Though we have proved that the proposed system presents some robustness against common manipulations such as re-coding at lower bit rates and frame dropping, we do not think it is a good candidate for copyright protection and Digital Rights Management applications, since the level of security required by these scenarios is far from being reached. In spite of this, the proposed algorithm stands out for its simplicity, flexibility and low computational burden, thus being a suitable candidate for a number of novel and interesting applications, such as content identification, enhanced services, legacy systems, conditional access, authentication and integrity verification.

Appendix

In this Appendix, the mean and variance of the accumulator Acc are derived by assuming that the video sequence is not watermarked.

If the coefficients composing each pair are sufficiently close each other, thus allowing us to assume that they are identically distributed, it is expected that

$$E[|QF_{i_1}|] = E[|QF_{i_2}|] \tag{11}$$

whereby $\{QF_{i_1}, QF_{i_2}\}$ the i^{th} of the N_p coefficients pairs that are used to carry the watermark is indicated. From the previous relation it derives that the expected value of Acc under hypothesis A, is:

$$\mu_{Acc|A} = E[Acc|A] = \sum_{i=1}^{N_p}(E[|QF_{i_1}|] - E[|QF_{i_2}|]) = 0. \tag{12}$$

For the variance, we have

$$\begin{aligned}
\sigma^2_{Acc|A} &= E[Acc^2|A] = \\
&= E\left[\sum_{i=1}^{N_p}(|QF_{i_1}| - |QF_{i_2}|)\sum_{j=1}^{N_p}(|QF_{j_1}| - |QF_{j_2}|)\right] \\
&= \sum_{i,j=1}^{N_p}(E[|QF_{i_1}||QF_{j_1}|] + E[|QF_{i_2}||QF_{j_2}|] \\
&\quad - E[|QF_{i_1}||QF_{j_2}|] - E[|QF_{i_2}||QF_{j_1}|]) \\
&= \sum_{i=1}^{N_p} E[QF_{i_1}^2] + \sum_{i\neq j}^{N_p} E[|QF_{i_1}||QF_{j_1}|] \\
&\quad + \sum_{i=1}^{N_p} E[QF_{i_2}^2] + \sum_{i\neq j}^{N_p} E[|QF_{i_2}||QF_{j_2}|] \\
&\quad - \sum_{i=1}^{N_p} E[|QF_{i_1}||QF_{i_2}|] - \sum_{i\neq j}^{N_p} E[|QF_{i_1}||QF_{j_2}|] \\
&\quad - \sum_{i=1}^{N_p} E[|QF_{i_1}||QF_{i_2}|] - \sum_{i\neq j}^{N_p} E[|QF_{i_2}||QF_{j_1}|] \\
&= \sum_{i=1}^{N_p} E[QF_{i_1}^2] + \sum_{i\neq j}^{N_p} E[|QF_{i_1}||QF_{j_1}|] \\
&\quad + \sum_{i=1}^{N_p} E[QF_{i_2}^2] + \sum_{i\neq j}^{N_p} E[|QF_{i_2}||QF_{j_2}|] \\
&\quad - 2\sum_{i=1}^{N_p} E[|QF_{i_1}||QF_{i_2}|] - \sum_{i\neq j}^{N_p} E[|QF_{i_1}||QF_{j_2}|] \\
&\quad - \sum_{i\neq j}^{N_p} E[|QF_{i_2}||QF_{j_1}|].
\end{aligned} \tag{13}$$

Given that $|QF_{i_1}|$ and $|QF_{i_2}|$ are independent for $i \neq j$, the following relation holds:

$$\begin{aligned}\sigma^2_{Acc|A} = &\sum_{i=1}^{N_p} E[QF^2_{i_1}] + \sum_{i\neq j}^{N_p} E[|QF_{i_1}|]E[|QF_{j_1}|]\\ &+ \sum_{i=1}^{N_p} E[QF^2_{i_2}] + \sum_{i\neq j}^{N_p} E[|QF_{i_2}|]E[|QF_{j_2}|]\\ &-2\sum_{i=1}^{N_p} E[|QF_{i_1}||QF_{i_2}|] - \sum_{i\neq j}^{N_p} E[|QF_{i_1}|]E[|QF_{j_2}|]\\ &-\sum_{i\neq j}^{N_p} E[|QF_{i_2}|]E[|QF_{j_1}|] \qquad (14)\end{aligned}$$

which, by considering (11) and by further assuming that $E[QF^2_{i_1}] = E[QF^2_{i_2}]$, becomes

$$\begin{aligned}\sigma^2_{Acc|A} &= 2\sum_{i=1}^{N_p} E[QF^2_{i_1}] + 2\sum_{i\neq j}^{N_p} E[|QF_{i_1}|]E[|QF_{j_1}|]\\ &\quad -2\sum_{i=1}^{N_p} E[|QF_{i_1}||QF_{i_2}|] - 2\sum_{i\neq j}^{N_p} E[|QF_{i_1}|]E[|QF_{j_1}|]\\ &= 2\sum_{i=1}^{N_p} E[QF^2_{i_1}] - 2\sum_{i=1}^{N_p} E[|QF_{i_1}||QF_{i_2}|]. \qquad (15)\end{aligned}$$

In conclusion, the expression of the variance under hypothesis A can be estimated as

$$\sigma^2_{Acc|A} \cong 2\left(A^2 - B^2\right) \qquad (16)$$

where

$$A^2 = \sum_{i=1}^{N_p} QF^2_{i_1} \cong \sum_{i=1}^{N_p} E[QF^2_{i_1}] \qquad (17)$$

and

$$B^2 = \sum_{i=1}^{N_p} |QF_{i_1}||QF_{i_2}| \cong \sum_{i=1}^{N_p} E[|QF_{i_1}||QF_{i_2}|] \qquad (18)$$

are estimated over the set of non watermarked MBs.

REFERENCES

[1] "TR 18 034-1 – Information Technology – Multimedia Framework (MPEG21) – Part 1," ISO, ISO/IEC JTC1/SC29/WG11, 2000.

[2] "N5229, Requirements for the Persistent Association of Identification and Description of Digital Items," ISO, ISO/IEC JTC1/SC29/WG11, 2000.

[3] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 19–31, Feb. 1997.

[4] E. Koch and J. Zhao, "Toward robust and hidden image copyright labeling," in *Proc. IEEE Workshop on Nonlinear Signal and Image Processing '95*, Neos Marmaras, Greece, Jun. 20–22, 1995.

[5] C. Hsu and J. Wu, "Hidden signatures in images," in *Proc. IEEE Int. Conf. Image Processing (ICIP'96)*, vol. 3, Sep. 1996, pp. 223–226.

[6] F. Hartung and B. Girod, "Digital watermarking of raw and compressed video," in *Proc. SPIE 2952: Digital Compression Technologies and Systems for Video Communication*, Berlin, Germany, Oct. 1996, pp. 205–213.

[7] B. G. Mobasseri, "Exploring CDMA for watermarking of digital video," *Proc. SPIE, Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 96–102, Jan. 1999.

[8] C. Hsu and J. Wu, "Digital watermarking for video," in *Proc. IEEE Int. Conf. Digital Signal Processing*, vol. 1, Jul. 1997, pp. 217–220.

[9] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 540–550, May 1998.

[10] T. Kalker, G. Depovere, J. Haitsma, and M. Maes, "A video watermarking system for broadcast monitoring," *Proc. SPIE, Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 103–112, Jan. 1999.

[11] F. Deguillaume, G. Csurca, J. O'Ruanaidh, and T. Pun, "Robust 3D DFT video watermarking," *Proc. SPIE, Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 113–124, Jan. 1999.

[12] F. Hartung and B. Girod, "Digital watermarking of MPEG-2 coded video in the bitstream domain," in *Proc. ICASSP'97*, vol. 4, Münich, Germany, April 1997, pp. 2621–2624.

[13] G. C. Langelaar, R. L. Lagendijk, and J. Biemond, "Watermarking by DCT coefficient removal: A statistical approach to optimal parameter settings," *Proc. SPIE, Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 2–13, Jan. 1999.

[14] G. C. Langelaar and R. L. Lagendijk, "Optimal differential energy watermarking of DCT encoded images and video," *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 148–158, Jan. 2001.

[15] A. Piva, R. Caldelli, and A. De Rosa, "A DWT-based watermarking system for MPEG-4 video streams," in *Proc. ICIP 2000, IEEE Int. Conf. Image Processing*, Vancouver, BC, Canada, Sep. 10–13, 2000.

[16] M. Barni, F. Bartolini, V. Cappellini, A. Lippi, and A. Piva, "DWT-based technique for spatio-frequency masking of digital signatures," in *Proc. SPIE, Security and Watermarking of Multimedia Contents*, vol. 3657, Jan. 1999, pp. 31–39.

[17] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Trans. Image Process.*, vol. 10, pp. 755–766, May 2001.

[18] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "Application-driven requirements for digital watermarking technology," in *Proc. EMMSEC98, Eur. Multimedia Microprocess. System and Electronic Commerce Conf. and Exhibition*, Bordeaux, France, Sep. 1998, pp. 513–520.

[19] J. Dittman, M. Stabenau, and R. Steinmetz, "Robust MPEG video watermarking technologies," in *Proc. 6th ACM Int. Conf. Multimedia*, Bristol, U.K., Sep. 1998, pp. 71–80.

[20] S. Lin and D. J. Costello, Jr, *Error Control Coding: Fundamentals and Applications*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[21] J. R. Hernandez, J. F. Delaigle, and B. M. M. Macq, "Improving data hiding by using convolutionale codes and soft-decision decoding," *Proc. SPIE, Security and Watermarking of Multimedia Contents II*, pp. 24–47, Jan. 2000.

[22] T. Kalker, "Digital video watermarking," in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, Florence, Italy, Jun. 1999.

Mauro Barni (M'96) graduated in electronic engineering in 1991 and received the Ph.D. degree in informatics and telecommunications in October 1995, both from the University of Florence, Florence, Italy.

From 1991 through 1998, he was with the Department of Electronic Engineering, University of Florence, where he was a Post-doctoral Researcher. Since September 1998, he has been with the Department of Information Engineering, University of Siena, Siena, Italy, where he is an Associate Professor. His main interests are in the field of digital image processing. His research activity is focused on the application of image processing techniques to copyright protection and authentication of multimedia data (digital watermarking), and to the transmission of image and video signals in error-prone, wireless environments. He is author/co-author of more than 130 papers published in international journals and conference proceedings, and holds three European patents in this field. He is on the editorial board of the *EURASIP Journal of Applied Signal Processing*.

Prof. Barni serves as associate editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE SIGNAL PROCESSING MAGAZINE (column and forum section). He is a member of the IEEE Multimedia Signal Processing Technical Committee (MMSP-TC).

Franco Bartolini (M'96) was born in Rome, Italy, in 1965. In 1991, he graduated (cum laude) in electronic engineering from the University of Florence, Florence, Italy. In November 1996, he received the Ph.D degree in informatics and telecommunications from the University of Florence.

Since November 2001, he has been an Assistant Professor at the University of Florence. His research interests include digital image sequence processing, still and moving image compression, nonlinear filtering techniques, image protection and authentication (watermarking), image processing applications for the cultural heritage field; signal compression by neural networks, and secure communication protocols. He has published more than 130 papers on these topics in international journals and conferences. He holds three Italian and one European patent in the field of digital watermarking.

Dr. Bartolini passed away on January 1, 2004.

Nicola Checcacci was born in Arezzo, Italy, in 1971. In April 1999, he graduated in electronic engineering from the University of Florence, Florence, Italy.

After a short cooperation in the field of image protection and authentication (watermarking) with the University of Florence, in April 2000 he joined the Network Services Division of Alcatel in Florence, working at the engineering and deployment of several WANs in Europe. In September 2000, he left Alcatel to join the Marketing department of TeleMedia International (TMI), a global operator that provides value added TLC services and customized solutions to multinational customers. In December 2001, he joined the marketing department of Telecom Italia. Since January 2003, he is Product Manager in the Marketing Department of Telecom Italia Sparkle a fully owned subsidiary of Telecom Italia. He cooperates in defining, implementing and developing the company's retail service portfolio.