Time-Variant Watermarking of MPEG-Compressed Digital Videos

Ernst L. Leiss*
Department of Computer Science
University of Houston
Houston Texas 77204-3010

coscel@cs.uh.edu
voice 713-743-3359
fax 713-743-3335

Abstract

Watermarks allow one to embed information into digital videos in a way that is imperceptible to the viewer. This information can be used to establish ownership, trace origin of copies, and verify the integrity of the video. Watermarking may be compared to injecting additional energy; to ensure that this injection remains unnoticeable, it should be as small as possible. We outline an approach that permits a significant increase of the amount of information that can be accommodated in a watermark without any increase in the complexity of the process, namely time-variant watermarks. Since data compression is an important aspect in storing and distributing digital videos, we formulate our approach assuming the video is represented in an MPEG format. We discuss implementation issues of time-variant watermarks, with special emphasis on their advantages over the usual time-invariant watermarks. We comment on defeating attacks using filtering, cropping, resizing, and other standard methods used to defeat watermarks, such as changing existing frames, as well as new attacks, such as removing, repeating or permuting frames.

Keywords: Digital video, MPEG compression, watermarks, time-variance.

* Support of this work under NSF Grant SFS-0313880 is acknowledged.

1. Introduction and Motivation

All digital information can be copied perfectly since any string consisting of 0's and 1's is indistinguishable from its copy [Leiss, 1982]. The existence of perfect copies has numerous implications for data security and integrity; note that in the physical world, perfect copies do not exist by definition. Since it is impossible to distinguish a copy from the original, if information is used to control access to resources, anyone who is able to copy the information will have access to those resources. Consequently, it is difficult to establish ownership of digital intellectual property: two parties may each claim to be the legitimate owner of digital information.

Digital watermarks are an attempt to address the problem of perfect copies in digital data. While they are not foolproof, they are a workable approach provided a few conditions are satisfied. Briefly, when using a digital watermark additional information is embedded into or superimposed on the original images. As concerns about establishing ownership of digital media have escalated in recent years (witness the claims by the recording industry blaming reduced sales of CDs on illicit file sharing), watermarks have attracted increased attention.

We may differentiate visible and invisible watermarks. Visible watermarks are often used in TV transmissions, where in a fixed location in each image or frame, a small logo identifying the transmitter is inserted, obliterating or obscuring that part of the image. Another type of visible watermark is provided by IBM's project watermarking a portion of the Vatican Libraries' holdings of images. Visible watermarks can typically be removed quite easily, thereby removing (a portion of) the information contained in the watermark. Of course, this creates the problem what to put in the place of the removed visible watermark if the watermark replaced it. Since it is not possible to restore the original image, an "empty" spot is left in the resulting image which could be filled by interpolation but this will frequently provide unsatisfactory results since it will still be possible to discern the (rough) shape of the removed logo, even though information contained in the logo would no longer be accessible. If the watermark was added instead, subtracting it restores the original image. For these reasons, invisible watermarks are preferred. Invisible watermarks change certain characteristics of the image, but this is done in a way that is not noticeable to the naked human eye. Here, we will consider exclusively invisible watermarks.

Depending on one's objectives, either robust or fragile watermarks can be used. Fragile watermarks have been proposed with the intention of degrading the watermark with each subsequent copy operation; thus, fragile watermarks are designed to limit the number of times a document may be passed on. Robust watermarks are of interest if one wishes to attach an indelible stamp of ownership; clearly the methods employed must be robust, that is impervious to various operations, such as rescaling, filtering, or superimposing an additional watermark. A variety of schemes designed to achieve these objectives have been proposed; see for example [Tanaka 1990, Matsui 1994, Bender 1995, Berghel 1997, Cox 1997, Barni 1998, Duan 1998, Lee 1999]. While robust and fragile watermarks can be considered complementary, it is the robust ones that serve in establishing ownership. In this paper, we will consider exclusively robust watermarks.

One aspect that has received little attention relates to the amount of information that can be encoded in a watermark. Clearly, robustness is directly correlated with the redundancy of the watermark; for example, if a certain small pattern is repeated many times in a watermark, the removal of the watermark through cropping an image is foiled. Similarly, the invisibility of a watermark is related to the extent of changes in the information that makes up the media; clearly, extensive changes will have a greater impact on the watermarked medium than small ones. In general, it is useful to view

the process of watermarking an image akin to injecting energy – the more energy is injected, the more the original image is impacted, to the point where this process may be perceptible to the viewer. This is of course the antithesis of invisibility! Thus, there are certain limits on the amount of information that can be encoded in the watermark. To alleviate problems created by the paucity of information available in the watermark, we propose the notion of time-variant watermarks. In this scheme, different frames of a video (or an audio) file will be associated with different watermarks. There are two advantages to this approach: it makes it much more difficult to defeat the watermark, and it lets one encode significantly more information in the watermark while permitting a great deal of redundancy and repetition. The most important practical aspect of this new scheme is that its computational complexity is identical to that of conventional watermarking: It is immaterial whether we embed the same watermark into many frames of the video or whether we embed a different watermark in each of these frames. It will become quite clear that the additional information in the watermark can be exploited to achieve increased protection of intellectual property.

All existing watermark schemes, visible and invisible, robust and fragile, are time-invariant: the embedded watermark is the same, independent of the video frame into which it is embedded. In contrast, our time-variant watermark scheme permits the embedding of sequences of watermarks into the medium to be watermarked. Typically, the watermark will consist of a number of frames that may be smaller than the number of frames of the medium into which the watermark is embedded. If it is smaller, then as in time-invariant watermark schemes, the watermark sequence is repeated until the end of the medium into which it is embedded is reached. This allows one either to increase the amount of information that is encoded in the aggregate watermark or to reduce the amount of information that is contained in a single watermark frame. Most interesting is the case where the number of watermark frames is equal to the number of frames of the file to be watermarked; here, the watermark can be used to insert sequencing information that is invisible to the viewer. This information can be used to detect, and demonstrate if required, whether original frames have been removed or permuted or repeated. In particular, the removal and the permutation of frames existing in a digital video cannot be detected using conventional, time-invariant watermarks: removing watermarked frames is cannot be detected, since everything remaining is properly watermarked, and the same argument applies to permuting existing frames. Note that inserting new frames is detectible using time-invariant watermarks since the new frames would not be watermarked.

We sketch our approach to time-variant, invisible, robust waterworks using video media; an analogous approach can be formulated for audio or similar media that represent information where a certain amount of errors can be tolerated. It is for example clear that viewers of television are perfectly willing to tolerate a fairly high percentage of “wrong” pixels, perhaps as high as 5% without major deterioration in the perceived quality of the image viewed. For audio the percentage may be somewhat lower, but still significant. On the other hand, an error rate of even 0.5% in text data would be considered quite unacceptable, as it amounts to about one error each three lines of text. This would be even more unacceptable if such “errors” were not randomly distributed, but instead deliberately introduced; here, even a single error (change of one character) might be devastating – consider for example a contract obliging A to pay US$10,000 and assume that the numeral “1” were changed to the numeral “9”!

An important aspect in storing and disseminating digital videos is the amount of data required to represent them faithfully. Clearly, we want to keep the file size as small as possible. Attempts to reduce the size of a video lead naturally to data compression techniques. Following industry standards, we assume JPEG encoding for individual (still) images as well as intracoded frames. We assume a standard MPEG organization of the video sequence into I (intracoded), P (predictive-

coded), and B (bi-directionally predictive-coded) frames. We outline our approach's advantages, in particular increased imperviousness against a variety of attempts to defeat the watermarking process, through filtering, cropping, resizing, and other operations, and quantify the increase in information content that can be accommodated in the new watermark.

We note that some of the objectives one pursues in using watermarks can be attained by other means, primarily encryption-based approaches [Leiss 1982]. For a discussion of these, see for example [Chen 1995, Chen 1996]. The current work is primarily based on research reported in two M. S. theses [Yang 1999, Yang 2001]. We will give a very brief review of goals and objectives in using watermarks. Then we give a sketch of the MPEG organization of a video file, with some attention paid to the representation of color and the JPEG technique for still images. Then we discuss time-variant watermarks in more detail and indicate the benefits obtained in this way. We conclude with a summary of the advantages of the approach and by indicating possible future work.

2. Watermark Goals and Objectives

We briefly review aspects of watermarks pertinent to our work [Bush 1999, Chun 1998, Cox 1997, Hsu 1998, Hsu 1999, Koch 1995]. The overall objective is the protection of intellectual property [Berghel 1997], in our case, the intellectual property contained in a digital video file.

As already mentioned, we are interested in invisible, robust watermarks. Robustness means that the watermark must be impervious to attempts at removing, destroying, obliterating, or overwriting it. Any attempt to do so should result in a very noticeable degradation of the image before the watermark is lost. Given the environment in which the watermark is used, the process of embedding the watermark must be compatible with MPEG processes. Consequently, watermarks must be able to survive both loss-less and lossy compression techniques, as well as other common video processing techniques, such as scaling, cropping, resizing, and filtering (in the case of color video, this includes changes in the color scheme, such as reducing the color palette [e. g., from 16 bit to 8 bit]).

The watermark must allow the legitimate owner of the video to demonstrate this ownership conclusively (for example, to a judge or adjudicator). Therefore, sufficient information be present that can be used for this purpose. Below we will argue that none of the existing, time-invariant watermark schemes fully attains this goal. The principal reason for this is the fact that within the context of MPEG-based compression, it is virtually impossible to guarantee that entire scenes have not been removed from the video nor that original scenes have been permuted or repeated. More specifically, a new scene in a video will almost certainly result in the use of an I-picture for the first frame of the scene. Since most of the watermark insertion concentrates on I-pictures, this implies that the removal of a group of frames (for example, an entire scene) that begins with an I-picture would not be noticeable if all watermark images are identical, that is, if they are time-invariant. It is true that the running time of a video could be used to **detect** this type of tampering, but it would not allow one to determine **where** the tampering occurred – for this, our time-variant watermark approach described in this paper is required.

Finally, we mention three important practical aspects of any watermark; failure to satisfy either one of them will render the approach unacceptable in practice:

1. The insertion of the watermark must not affect the perceived quality of the video. While the watermark information is of course embedded in the signal (i. e., the original signal is modified by the watermark insertion), this must not affect the **perceived** quality of the signal.

2. The process of inserting the watermark must not substantially increase the overall complexity of generating and using the video, at least not significantly beyond what MPEG already requires. This is the reason why certain cryptography-based signature schemes (see [Chen 1995, Chen 1996]) are not acceptable in practice, even though they could be made arbitrarily secure.

Clearly, if the quality of the video images is visibly affected by the watermark insertion, viewers will refuse to accept the resulting lower quality of the images. On the other hand, no matter how well the image quality is preserved, if the process of inserting the watermark adds a significant amount of processing to the already somewhat time-consuming MPEG processing requirements, there may simply not be sufficient compute power to carry out the watermark **insertion** in real time. Note that the complexity of **viewing** a watermarked video is not increased by the watermark, in contrast to encryption-based approaches.

3. It must be possible to demonstrate legally that only the true owner of a video is capable of embedding the watermark. If this is not possible in a legally binding way, the utility of a watermark for the protection of intellectual property is seriously compromised.

3. MPEG Compressing of Video Files

We sketch the organization of MPEG-compressed video files, starting with JPEG for still images which forms the basis of MPEG. First however we explain color representation, with its implications for the embedding of watermarks.

MPEG is essentially a (family of) method(s) for compressing a video file. An ordinary 24-bit image with 640*480 pixels requires almost 1MB of space (high-definition digital images would require even more). Since there are 30 frames per second in a typical digital video, a one-hour video amounts to about 100 GB of data. This amount of raw data contains a great deal of redundancy, the reduction of which is the goal of the use of (one of) the MPEG techniques. All MPEG schemes are based on the JPEG technique, applied to (some of the) individual frames of the video.

In virtually all videos, both individual frames themselves and the succession of frames contain much overt redundancy. For example, the background of a scene will ordinarily not change from one frame to the next unless the camera moves (temporal redundancy); moreover, this background may be virtually featureless and constitute a relatively large percentage of the frame (consider recording an interview), resulting in a large degree of redundancy (spatial redundancy).

Both the storage of the video file and the bandwidth requirements that result when transmitting digital video files over a network are a concern that data compression is designed to address. Were the above mentioned video file transmitted in its raw format, the minimum bandwidth necessary for sending a single digital video would be 240 Megabits per second (without leaving bandwidth for anything else). This is of course unmanageable (consider for example video-on-demand services).

JPEG [Wallace 1992] stands for Joint Photographic Experts Group, an ISO/CCITT committee and is a standardized compression technique for full-color or gray-scale images of realistic digital

images. (It is not designed for line drawings or lettering although the presence of such features is not an impediment for JPEG.) It is based on a lossy compression technique known as the Baseline method; this is a scheme that employs the DCT (Discrete Cosine Transform, see [Wallace 1992]). A compression technique is called **loss-less** if the information content of the original file can be retrieved from the compressed file in its entirety, without sacrificing accuracy or precision. A technique is called **lossy** if the compressed file loses some of the original information. Although a loss-less approach appears more attractive, it is typically the lossy techniques that result in significantly larger savings. Most importantly, the loss of information they suffer is typically imperceptible to the viewer. It is not unusual to obtain a compression ratio (uncompressed file size compared with compressed file size) of 15 or more with excellent image quality; this compression ratio can be even higher if some deterioration of the image quality is acceptable [Sonka 1998]. JPEG, and consequently MPEG, allows the user to specify the image quality in terms of rather intuitive parameters. In this way, the image quality can be varied, depending on the given application. For example, a major motion picture may be encoded with greater faithfulness (and at greater cost in storage space or transmission bandwidth) than a video conference in a corporate setting. JPEG is considered a very popular and efficient coding scheme for continuous-tone still images. It also forms the basis of the MPEG family of approaches to encoding digital video. Before we describe MPEG-2 (which is at present the main representative of the MPEG schemes applicable to digital video), we give some brief explanation of digital color and its representation.

Humans perceive colors as combinations of the primary colors red, blue, and yellow (the typical rainbow arrangement). Only slightly deviating from this, video hardware generally uses the RGB model (Red, Green, Blue) with a pixel being associated with a triple (RGB) representing the color intensities; (000) represents black in this scheme (absence of everything), (kkk) white (presence of everything), (k00) pure red, and so on, where the value k is the quantization granularity for each primary color (for a total of k+1 different values, namely 0 through k). Thus, if k is 255 (a very common choice since it amounts to one byte), there are 2^{8+8+8} or 2^{24} different representable colors. Clearly, smaller values of k correspond to less faithfulness in the color scheme, larger values to greater faithfulness. With few exceptions (a contrary example might be a fairly uniform sky that continuously goes from light blue to gray), color schemes with more than 24 bits (eight for each of the three primaries) result in improvements in image quality that are virtually imperceptible to the unassisted human eye.

In practical applications, the RGB signal is usually transformed into one that is displayable with fewer major artifacts on black-and-white devices (including printers!), namely the (Y, C_b, C_r) representation, where Y is the luminance, C_b is the blue chrominance, and C_r the red chrominance. (R, G, B) and (Y, C_b, C_r) correspond to each other linearly [Benoit 1997]:

$$Y = 0.587\,G + 0.299R + 0.114\,B$$
$$C_b = 0.564\,(B - Y)$$
$$C_r = 0.713\,(R - Y)$$

It is important for the design of data compression techniques to understand that the human eye is less perceptive for color than for luminance. This implies for natural images that the chrominance components of a signal can tolerate a more reduced bandwidth than the luminance component, without affecting significantly the perceived image quality. Typically, the bandwidth for chrominance may be chosen to be one half to one quarter of that for luminance without affecting human perception [Benoit 1997].

Lossy JPEG compression consists of six main steps [Wallace 1992]:

1. Decomposition of the image into blocks of size 8*8 pixels; each block can be viewed as a 64-point discrete signal which is a function of the two spatial dimensions.

2. The Discrete Cosine Transform is applied to each 8*8 matrix which generates a new 8*8 matrix consisting of the coefficients of increasing spatial frequency. These coefficients can be viewed as the relative amount of the 2D spatial frequencies in the 64-point input signal. The coefficient with frequency 0 in both dimensions is referred to as the DC coefficient while the other 63 are the AC coefficients.

3. Quantization (or discretization) is applied to the 64 DCT coefficients to yield an 8*8 Quantization table Q(u,v) consisting of integers. As result of the DCT operation, the values in Q increase from left to right and from top to bottom. This takes into account the peculiarities of human vision, in particular the fact that the human eye does not distinguish very fine details below a certain luminance level.

4. The 63 AC coefficients in Q are concatenated into a zigzag scan; in terms of (u,v), this scan is

DC: 00

AC: 01 10
20 11 02
03 12 21 30
40 31 22 13 04
05 14 23 32 41 50
60 51 42 33 24 15 06
07 16 25 34 43 52 61 70
71 62 53 44 35 26 17
27 36 45 54 63 72
73 64 55 46 37
47 56 65 74
75 66 57
67 76
77

This helps in entropy coding by placing low-frequency coefficients, which are more important in perception, before high-frequency ones.

5. Run-length coding replaces a sequence of identical values by one indication of that value followed by the number of these values in the sequence. This is where major compression in JPEG occurs, since from a certain point p on in the sequence of the 63 AC coefficients of the zigzag scan, we can replace the remainder by zeroes without affecting the visual quality of the image. The value of p is a parameter in this process: if p is small, say 5, the image quality is reduced and the compression greatly improved; if p is large, say 30, the image quality is virtually unaffected but at the cost of reduced compression. In fact, studies of human perception of typical images have shown that even for relatively small values of p, say around 10, the perceived quality of the image is virtually unaffected. While the value at which people will notice a difference depends on the type of image, it is a very important aspect of JPEG to determine as small a value of p as is acceptable from a visual perception point of view.

6. The final step consists of applying Huffman coding to the resulting sequences; this further reduces the amount of data to be transmitted.

MPEG is based on JPEG and is designed to remove temporal redundancies (redundancies that occur from one frame to the next) after JPEG has been applied to remove the spatial redundancies within each frame. Temporal redundancies are detected by motion estimation whereby portions of images in consecutive frames are matched up. Three fundamental types of pictures are distinguished in this process, namely I-pictures, P-pictures, and B-pictures. Intra or I-pictures are encoded without any

reference to other frames, while Predicted or P-pictures and Bi-directionally Predicted or B-pictures depend on other frames, for P-picture only on the preceding I- or P-picture, for B-picture on I- and P-pictures both preceding and following it. The number of P-pictures between two consecutive I-pictures is an important parameter: Since much redundancy is detected (and removed!) between I-pictures, making this value large results in more savings. However, making it too large will affect the quality of the interpolated image frames. B-pictures fill in the gaps between I- (and P-) pictures and provide the largest savings. The objective is to have as few I- (and P-) pictures as is possible without affecting the visual quality of the video. Since typically there are many more B-pictures than I- or P-pictures, ratios of 200 can be achieved in video compression without sacrificing a great deal of quality [Sonka 1998]. This value of 200 is the combination (product) of the JPEG compression ratio and compression ratio resulting from the removal of redundancy related to motion estimation, based on the I-, P-, and B-pictures. To give a few specific values, if M (N) is the number of pictures between two successive P-pictures (I-pictures), then typical values might (3,12). Thus, 1/12 of a group of pictures are I-pictures, 1/4 P-pictures, and 2/3 B-pictures.

Motion estimation involves defining a motion vector, which establishes the correlation between a "departure" zone in the first picture and an "arrival" zone in the second. This is done on the basis of macro blocks (blocks of size 16*16, or four 8*8 blocks of luminance, one 8*8 block for red chrominance, and one 8*8 block for blue chrominance). This allocation of four times the amount of data for luminance than for each of the chrominance values reflects the differing levels of perception of the naked human eye.

4. Time-Variant Watermarks

First, we briefly review the process of embedding a (time-invariant) watermark into an MPEG-2 digital video file. In essence, our approach is applicable to any watermarking scheme. Thus, we are less interested in a specific scheme; instead, we describe the differences between the traditional, time-invariant approach and our time-variant method, outline the advantages of our approach, and indicate how it can be used to attain higher levels of protection of intellectual property.

Numerous approaches to embedding (time-invariant) watermarks into images have been described in the literature. They can be grouped into two major categories, namely methods that embed the watermark by modifying directly the intensity of (some or all of) the pixels of an image [Bender 1995, Nikolaidis 1998], and methods that act upon (some or all of) the coefficients of an underlying transform domain (most common are the Discrete Cosine Transform or DCT and the Discrete Fourier Transform or DFT) [Koch 1994, Boland 1995, Cox 1997, Barni 1998, Duan 1998, Lee 1999]. While we concentrated in [Yang 1999, Yang 2001] on the method described in [Cox 1997], it should be clear that any of the domain-based approaches would do nicely in an MPEG environment. The underlying idea is the following notion known as spread spectrum technique: The frequency domain of the image to be watermarked is viewed as a communication channel and the watermark is viewed as a signal that is transmitted through it. Thus, the watermark is spread over many frequencies so that the energy change in any one frequency is small enough to render it imperceptible. The objective is of course that the embedded watermark survive common signal manipulations (such as lossy and loss-less compression, filtering, conversions between digital and analog representation) and geometric manipulations (such as cropping, scaling, translation, rotation). In addition to these, superposition of one or more additional watermarks should also be detectable. Finally, manipulations related to sequencing of pictures in a video are of concern; these include in particular adding new or removing original pictures. Another requirement relates to the

ability to demonstrate conclusively to a judge one's ownership of the original video, that is, the owner, and only the owner, should be able to do this. We refer to the literature for the technical details of inserting the (time-invariant) watermark. For our purposes, it suffices to note that techniques exist which meet the stated requirements and which are sufficiently simple and efficient to permit their implementation within the context of an MPEG-2 video file without increasing the complexity of the operations involved in generating, viewing (or possibly removing the watermark), or adjudicating a watermarked video [Busch 1999].

An important aspect of watermarking within an MPEG context is the determination which pictures of a video file are to be watermarked. On the basis of our brief description, it is clear that the watermarking process involves individual frames or pictures which are subjected to JPEG compression. This implies that the watermark should be inserted into the AC coefficients that occur quite early in the zigzag scan, since later AC coefficients may simply be removed (set to 0) without affecting the visual quality of the image. There are different techniques that ensure that the injection of energy (that is, the embedding of the watermark) into these coefficients does not distort their values unduly. As noted in [Benoit 1997, Katzenbeisser 2000], this approach is robust and affects the visual quality only minimally. Given the process of MPEG compression, we have three types of pictures, I-, P-, and B-pictures. Since only I-pictures are independently encoded in MPEG, watermark insertion concentrates on I-pictures. However, this does not imply that P- or B-pictures are unaffected, in as much as they depend on (watermarked) I-pictures (as they are interpolated based on these pictures) and thus are indirectly watermarked.

A watermark is typically an image that is substantially smaller that the video frame; for example, assuming a 640*480 pixel image, the watermark may be of size 80*60. Also, it may be black-and-white in order to reduce the amount of information that must be accommodated in each video picture. The watermark image would then be repeated 64 times to fill the entire image region.

Time-invariant watermarking schemes embed the same watermark picture into all pictures that are explicitly watermarked (essentially all the I-pictures). In contrast, our approach to providing time-variant marking schemes takes a watermark **video** consisting of a number N_0 of pictures and embeds this video in the usual way, frame by frame. Specifically, into the video frame number i to be watermarked we embed the watermark frame number i, for i=1, 2, …, N_0. If there are more than N_0 pictures in the video to be watermarked, the next batch of N_0 pictures get a second copy of the watermark video embedded, and so on, until the end of the video to be watermarked is reached. The number N_0 is a parameter: if N_0=1, then we have the ordinary, time-invariant watermarking scheme; if N_0 is greater than 1, the approach is time-variant. A sensible upper bound for N_0 is the number of I-pictures in the original video.

The information contained in the N_0 frames of the watermark is entirely up to the user. It is however useful to provide some sequencing information in the watermark video because if N_0 is equal to the total number of I-pictures in the video this will enable one to ensure that no original pictures had been removed from the watermarked video. Note that this is one operation that traditional, time-invariant methods are entirely unable to detect since the removal of an entire scene (starting with an I-picture) is undetectable. Other subversions that time-invariant watermarks are unable to detect, but that time-variant watermarking handles with ease, are permutations and repetitions of existing (that is, properly watermarked!) frames in the video.

We remark that in both time-invariant and time-variant watermarking, information is injected into the signal corresponding to each of the watermarked frames, the time-invariance of this information is very wasteful. In contrast, although our approach will inject no more energy into each of the

watermarked images than the traditional methods, the information content our approach allows us to embed is dramatically greater, since it changes from one watermark frame to the next. Furthermore, the time-complexity of inserting a time-invariant watermark is identical to that of inserting a time-variant watermark. Thus, time variance provides significantly greater functionality at no cost whatsoever!

The following table summarizes in what way an attack against a watermark is foiled; here INV indicates that the traditional time-invariant watermarking scheme will guard against this attack or manipulation (preserving the watermark) while VAR indicates that this is achieved by time-variant watermarks:

loss-less compression	INV, VAR
lossy compression	INV, VAR
filtering	INV, VAR
conversion (digital ↔ analog)	INV, VAR
cropping	INV, VAR
scaling	INV, VAR
translation	INV, VAR
rotation	INV, VAR
superposition of another watermark	INV, VAR
adding new frames	INV, VAR
removing original frames:	**VAR**
permuting original scenes:	**VAR**
repeating original scenes or frames:	**VAR**
legal demonstration of ownership:	INV, VAR

Recently, an experimental implementation of time-variant watermarks has been carried out by Ms. Enohi Ibekwe as part of a graduate project in Information Assurance. Specifically, she incorporated Koch and Zhao's method [Koch 1995] into MPEG-1 encoded video. This was done using the Dali multimedia library on a scripted language (TCL). Watermarks were inserted only into I-frames (which are compressed independently of any other frames), while the P- and B-frames were not watermarked; note however that vestiges of watermarks are present in these frames as well, since they depend on the (watermarked) I-frames. The results confirmed the observations above.

5. Conclusions

We have outlined our approach to embedding time-variant watermarks into digital video files. This method permits a significant increase of the amount of information over conventional, time-invariant watermarks while retaining all the advantages of conventional watermarking. In particular, the additional cost incurred by introducing time-variance is zero. The approach was formulated assuming the video file is represented in an MPEG-2 format, involving I-pictures, P-pictures, and B-pictures. In view of the standard data compression algorithm underlying MPEG-2, frames of the watermark video are embedded into the I-pictures, that is, those pictures of the video that are encoded independently, using JPEG techniques. We discussed implementation issues of time-variant watermarks, as well as their advantages over the usual time-invariant watermarks. In particular, this watermarking scheme permits one to defeat not just the usual attacks involving filtering, cropping, resizing, and changing color schemes, but also to guard against new attacks, namely removing or repeating frames as well as permuting scenes of the video. Important is that the complexity of the operations of embedding the watermark, viewing the watermarked video,

removing the watermark from the video, and the adjudication of the watermark remains unaffected by the watermark. Specifically, embedding of the watermark is incorporated in the MPEG-2 compression scheme and adds an insignificant amount of work; viewing the video is completely unaffected by the watermark, removing the watermark amount to MPEG-2 compression, and adjudication is essentially equivalent to extracting the watermark (which is in turn the same as removing it).

Bibliography

[Barni 1998] M. Barni, F. Bartolini, V. Cappellini, and A. Piva: A DCT-Domain System for Robust Image Watermarking, Signal Processing 66, 1998, 357-372.

[Bender 1995] W. Bender, D. Gruhl, and N. Morimoto: Techniques for Data Hiding, Proc. SPIE 2420, 1995, 164-173.

[Benoit 1997] H. Benoit: *Digital Television MPEG-1, MPEG-2 and Principles of the DVB System*, Arnold London, UK, 1997.

[Berghel 1997] H. Berghel and L. O'Gorman: Digital Watermark, http://www.acm.org/~hlb/publications/dig_wtr/dig_watr.html, 1997.

[Boland 1995] F. M. Boland, J. J. K. O Ruanaidh, and C. Dautzenberg: Watermarking Digital Images for Copyright Protection, Image Processing and Its Applications, 1995, 326-330.

[Bush 1999] C. Bush, W. Funk, and S. Wolthusen: Digital Watermarking Using DCT Domain Constraints, Proc. Int'l Conf. Image Processing, Vol. 3, 1996, 231-234.

[Chen 1995] F. Chen: Multimedia Authentication, M. S. Thesis, December 1995, Department of Computer Science, University of Houston.

[Chen 1996] F. Chen and E. L. Leiss: Authentication for Multimedia Documents, Proceedings, CLEI PANEL'96 - Conferencia Latinoamérica de Informática, June 3-7, 1996, Bogotá, Colombia, 613-624.

[Chun 1998] T. Chun, M. Hong, Y. Oh, D. Shin, and S. Park: Digital Watermarking for Copyright Protection of MPEG2 Compressed Video, IEEE Trans. Consumer Electronics, Vol. 44, No. 3, 1998.

[Cox 1997] I. Cox, J. Killian, T. Leighton, and T. Shamoon: Secure Spread Spectrum Watermarking for Multimedia, IEEE Trans. Image Processing, Vol. 6, No. 12, 1997, 1673-1687.

[Duan 1998] F. Y. Duan, I. King, L. W. Chan, and L. Xu: Intra-Block Algorithm for Digital Watermarking, IEEE Proc. Int'l Conf. Image Processing, Vol. 2, 1998, 1589-1591.

[Hsu 1998] C. Hsu and J. Wu: DCT-Based Watermark for Video, IEEE Trans. Consumer Electronics, Vol. 44, 1998, 206-216.

[Hsu 1999] C. Hsu and J. Wu: Hidden Digital Watermarks in Images, IEEE Trans. Image Processing, Vol. 8, No. 1, 1999, 58-68.

[Katzenbeisser 2000] S. Katzenbeisser and F. A. P. Petitcolas: Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, Boston, MA, 2000.

[Koch 1994] E. Koch, J. Rindfrey, and J. Zhao: Copyright Protection for Multimedia Data, IEEE Proc. Int'l Conf. Digital Media and Electronic Publishing, 1994, 203-213.

[Koch 1995] E. Koch and Z. Zhao: Toward Robust and Hidden Image Copyright Labeling, Proc. IEEE Workshop Nonlinear Signal and Image Processing, 1995, 443-449.

[Lee 1999] C.-H. Lee, H.-S. Oh, Y. Baek, and H.-K. Lee: Adaptive Digital Image Watermarking Using Variable Size of Blocks in Frequency Domain, Proc. IEEE Region 10 Conf. on TENCON, Vol. 1, 1999, 702-705.

[Leiss 1982] E. L. Leiss: *Principles of Data Security*, Plenum Publishing Corp., New York, NY, 1982.

[Matsui 1994] K. Matsui and K. Tanaka: Video Steganography, J. Interactive Multimedia Association Intellectual Property Project, Vol. 1, No. 1, 1994, 187-206.

[Nikolaidis 1998] N. Nikolaidis and I. Pitas: Robust Image Watermarking in the Spatial Domain, Signal Processing, Vol. 66, No. 3, 1998, 385-403.

[Sonka 1998] M. Sonka, V. Hlavac, and R. Boyle: *Image Processing, Analysis, and Machine Vision*, PWS, Pacific Grove, CA, 1999.

[Tanaka 1990] K. Tanaka, Y. Nakamura, and K. Matsui: Embedding Secret Information into a Dithered Multilevel Image, Proc. 1990 IEEE Military Communications Conf., 1990, 216-220.

[Wallace 1992] G. K. Wallace: The JPEG Still Picture Compressing Standard, IEEE Trans. Consumer Electronics, Vol. 38, No. 1, 1992, 18-35.

[Yang 1999] Z. Yang: Time-Variant Watermarks in Digital Video, M. S. Thesis, December 1999, Department of Computer Science, University of Houston.

[Yang 2001] Q. Yang: Time-Variant Watermarks in Color Digital Video, M. S. Thesis, December 2001, Department of Computer Science, University of Houston.