Perceptual Watermarks for Digital Images and Video

Raymond B. Wolfgang [†], Christine I. Podilchuk [‡], and Edward J. Delp [†]

[†] Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana
USA

[‡] Bell Laboratories
Lucent Technologies
Murray Hill, New Jersey
USA

Corresponding Author:

Edward J. Delp
School of Electrical and Computer Engineering
Purdue University
1285 Electrical Engineering Building
West Lafayette, IN 47907-1285
USA

Tel: +1 765 494 1740
Fax: +1 765 494 0880

Email: ace@ecn.purdue.edu

ABSTRACT

The growth of new imaging technologies has created a need for techniques that can be used for copyright protection of digital images. Copyright protection involves the authentication of image content and/or ownership and can be used to identify illegal copies of a (possibly forged) image. One approach for copyright protection is to introduce an invisible signal known as a digital watermark in the image.

In this paper, we describe digital image watermarking techniques, known as perceptually based watermarks, that are designed to exploit aspects of the human visual system. In the most general sense, any watermarking technique that attempts to incorporate an invisible mark into an image is perceptually based. However, in order to provide transparency (invisibility of the watermark) *and* robustness to attack, more sophisticated use of perceptual information in the watermarking process is required. Several techniques have been introduced that incorporate a simple visual model in the marking procedure. Such techniques usually take advantage of frequency selectivity and weighing to provide some perceptual criteria in the watermarking process. Even more elaborate visual models are used to develop schemes that not only take advantage of frequency characteristics but also adapt to the local image characteristics, providing extremely robust as well as transparent schemes. We present examples from each category - from the simple schemes that guarantee transparency to the more elaborate schemes that use visual models to provide robustness as well as transparency.

1   INTRODUCTION

In the last five years there has been an explosion in the use of digital imaging. Digital images are now widely distributed on the Internet and via CD-ROM. Digital imaging allows an unlimited number of copies of an "original" to be easily distributed and/or forged. This presents problems if the image is copyrighted. The protection and enforcement of intellectual property rights has become an important issue in the "digital world." Many approaches are available for protecting digital data; traditional methods include encryption, authentication and time stamping. In this paper we present algorithms for image authentication and forgery prevention known as *digital watermarks*.

This paper focuses on invisible watermarks that are designed to exploit perceptual information in the watermarking process. These watermarks are known as *perceptual watermarks*. We will first describe the problem, purpose and requirements of digital watermarking. An overview of visual models and their development is then presented, followed by some examples of image independent perceptual watermarks. Image dependent perceptual watermarks will then be examined, including transform-based algorithms. The emphasis in this paper is the class of perceptual techniques known as *image adaptive watermarks*; these are examined in detail. Several video watermarks and their unique requirements are then presented.

This paper describes image watermarking from a technical perspective. It is important to note that any technique that allows a user to assert their ownership of any digital object must also be placed in the context of intellectual property right law. In the final analysis how "well" a watermarking technique works depends on how effectively the technique protects an owner's intellectual property rights in a court of law. Overviews of digital watermarking are presented in [1,2,3] and in other papers in this special issue, particularly [4,5].

1.1   What is the Watermarking Problem?

Digital watermarking aids owners in asserting their intellectual property rights on the works of art they create. These rights are particularly difficult to enforce with digital images, since it is easy to copy and distribute perfect copies of an original image. Figure 1 shows that the basic components of any watermarking technique consist of a marking algorithm that inserts information, the watermark, into an image. The watermark is inserted into the image in the spatial domain or spatial frequency domain. As part of the watermarking technique, a testing algorithm must be defined that tests an image to see if a particular watermark is contained in the image. It is also desirable for the testing procedure to determine if the image has been altered and to supply localization information as to where the image was altered. It is our feeling that to assert ownership that is consistent with current intellectual property right law, the watermarking technique must support the use of third-party cryptographic-based digital time stamping that is embedded in the image through the watermarking process [6]. In [7] several general scenarios were identified where copyright enforcement is needed:

1. *Invisible watermarking for content and/or author authentication*
   An example of this scenario is images taken by a digital camera used in news gathering. The images must be watermarked upon capture so that the news services can be sure that an image has not been altered. The unrestricted distribution of copies of the images is

much less a concern here than verifying an image's origin and content. This is a very critical issue in the protection of historical images and images used in courts of law as evidence [8].

2. *Invisible watermarking for detecting unauthorized copies of images*
    Such watermarks can be used as follows:
    - To prove ownership of an image. The mere presence of the owner's mark in a suspect image can prove the theft of that image by the unauthorized possessor.
    - To identify the customer (who is a potential illegal distributor). The mark would represent the original purchaser whose copy has been illegally distributed.

Requirements and properties for the digital watermarks in the scenarios above differ. For authentication, it is important that even slight changes to the image be detected and localized. Embedding a false mark must be practically impossible and the watermark must be easily destroyed. It is not desirable for the watermark to remain in the image after attacks on the image such as filtering, although the watermark should survive cropping. These types of watermarks are known as *fragile watermarks*.
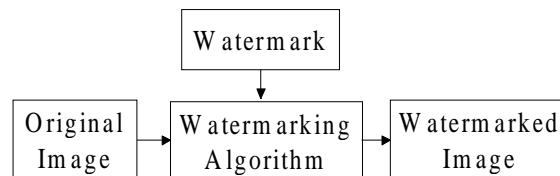
Figure 1. Block diagram of a watermarking algorithm

When an image is marked using a fragile watermark, an attacker does *not* want changes that they make to the image to alter the mark; it is desired by the attacker to have an altered image "pass" as authentic. This is not the case in the second scenario. Here, an attacker wants to remove the watermark at a minimal loss in image quality. In this way the true owner cannot verify the presence of the watermark in the image, thus greatly reducing any ownership claim. The watermark, therefore, must remain in the image after many types of attacks. These attacks include compression to low data rates, filtering, printing and rescanning, and geometric attacks such as cropping, resampling and rotation. Furthermore, users must not be able to attack the watermark through collusion by comparing multiple copies of the image marked with different watermarks. These watermarks are known as *robust watermarks*.

In this paper we will describe invisible watermarks that are designed to use perceptual information based on human visual system models. There are three main principles that characterize perceptually based watermarks:

1. Transparency – the watermark is not visible in the image under typical viewing conditions.
2. Robustness to attacks – the watermark can still be detected after the image has undergone linear or nonlinear operations such as those mentioned above.

3. Capacity – the watermarking technique must be capable of allowing multiple watermarks to be inserted in an image, with each watermark still being independently verifiable.

Additional tools are available to protect images from unauthorized viewing and modification. These include encryption techniques and digital time stamping [6,9]. Encryption disguises the content of an image and makes it "unreadable" to most applications. Only users who possess the decryption "key" can convert the encrypted image data back to its original form. Asymmetric encryption (i.e. public key cryptography) can be used for the authentication of a digital image. Authentication does not protect the content of the image, but proves who created the image. We feel that the use of "pure" encryption and authentication has limited use in the imaging area because when these techniques are used an image no longer is an image that can be viewed. Time stamping pinpoints the image owner *and* the time at which the image was generated. One trivial way to prove ownership of an image is to own the earliest reliable time stamp of the image. We believe that time stamping is absolutely critical to the success of any multimedia security system. We feel that all image watermarking techniques must use some form of third-party cryptographic time stamping to be successful. In many cases the time stamp can be used as part of the watermark generation. Time stamps also thwart the "re-watermarking attack" described in [10].

For a watermarking technique to be successful it must be part of a larger media protection system. For this reason copyright protection schemes should use various security tools in addition to digital watermarking. Such systems include the Copysight system from Intellectual Protocols 2 [11] and the System for Copyright Protection (SysCoP) [12].

Applications that will require image watermarking include: Internet imaging, digital libraries, digital cameras, medical imaging, image and video databases, surveillance imaging, video-on-demand systems, and satellite-delivered video. Several commercial companies exist that are selling watermarking software including Digimarc, IBM, and Signum Technologies.

1.2   Why use Visual Models?

The simplest form of a perceptually based watermarking scheme relies on incorporating the watermark into the perceptually insignificant parts of an image in order to guarantee transparency. An example of such a scheme is to make sure that the watermark signal amplitude is quite low by inserting information only to the low order bit planes in an image. However, watermarks which are embedded into the insignificant portions of an image are easily removed or altered through mild filtering or random bit-flipping in the lower bit planes without affecting the original image quality. If the amplitude of such watermarking schemes is increased to thwart these attacks and make the scheme more robust, the mark may become visible. For this reason, researchers have examined using more sophisticated ways of incorporating perceptual knowledge into watermarking schemes in order to provide robustness as well as transparency.

Several very effective frequency domain techniques have been introduced which take advantage of some type of frequency weighing in order to provide robust watermarking schemes. These techniques range from simple frequency selection based on common sense rules to frequency weighing based on models of the human visual system. These techniques may be image independent or dependent but usually do not take full advantage of the local image

characteristics in computing the watermark. Image-adaptive techniques not only take advantage of general frequency sensitivity but also rely on adapting the watermark to local image properties in order to provide maximum performance in terms of robustness while maintaining the transparency constraint.

We begin by introducing some visual models that have been developed in the context of still image compression in order to provide perceptually based quantization step sizes or bit allocation. Such models are ideally suited for the watermarking problem since thresholds developed to determine the visibility of quantization noise can be used to determine the perceptual upper bound on the watermark signal strength. This section is followed by a general review of watermarking techniques for still images. The review starts with the simplest techniques, motivated by the transparency requirement, to more elaborate frequency-domain techniques motivated by robustness as well as transparency. It continues with image-adaptive techniques, which provide the optimum performance in terms of robustness and transparency. We then describe how some of the perceptually based still image watermarking techniques have been extended to video watermarking.

2   VISUAL MODELS - OVERVIEW

2.1   Review of Perceptual Models

There has been much work over the years on trying to understand the human visual system as well as applying this knowledge to different image and video applications. In particular, a very useful application for perceptual models is in the area of source coding or compression. While traditional coding techniques take advantage of signal statistics to remove redundancy, it is ultimately the viewer who decides how well the compressed version represents the original source. Perceptual models allow us to take advantage of characteristics of the human visual system in order to remove irrelevancy as well as redundancy in designing the optimal compression algorithm.

2.1.1   Just-noticeable difference thresholds

Here we describe three different properties of the human visual system that have been studied in the context of image coding: *frequency sensitivity*, *luminance sensitivity* and *contrast masking*. Most of the early work on perceptually based image coding has utilized the frequency sensitivity of the human visual system as described by the *modulation transfer function* (MTF) in the coder design. This function describes the human eye's sensitivity to sine wave gratings at various frequencies. From such a model, given that the minimum viewing distance is fixed, it is possible to determine a static *just noticeable difference* (JND) threshold for each frequency band. These thresholds can be used for both quantization and bit allocation. Frequency sensitivity provides a basic visual model that depends only on viewing conditions and is independent of image content.

A further refinement can be achieved by extending the visual model to include luminance sensitivity. Luminance sensitivity is a way to measure the effect of the detectability threshold of noise on a constant background. This phenomenon depends on the average luminance value of the background as well as the luminance level of the noise. For the human visual system, this is a nonlinear function. Since luminance sensitivity takes advantage of local luminance levels it is

important that the frequency decomposition chosen for the coder allows for some local spatial control.

Frequency sensitivity and luminance sensitivity are good starting points in utilizing properties of the human visual system for image compression. However, frequency sensitivity depends only on viewing conditions; also, luminance sensitivity is a conservative estimate of visual masking which does not model masking properties due to high frequency details or texture. Ideally we would like a more dynamic model that allows for finer control of the quantization process. The addition of contrast masking allows for even more dynamic control of the JND threshold levels. Contrast masking refers to the detectability of one signal in the presence of another signal; the effect is strongest when both signals are of the same spatial frequency, orientation and location [13,14]. The most effective visual models should take into account frequency sensitivity, local luminance sensitivity and contrast masking.

The choice of the analysis filter bank used in a coder can affect the performance of the compression system and how effectively visual masking can be utilized. In order to utilize visual properties effectively, the filter bank should be selected to allow control of the spatial frequency location of the quantization distortion. Ideally, the addition of quantization distortion to one coefficient should not show up in coefficients that are not adjacent to the one that was perturbed. The analysis filter bank should also mimic the human visual system's structure in order to gain the most in terms of masking. For a human, this structure is a set of filters with frequency spacing of approximately 1.5 octaves and an angular width of 40 degrees [13].

The Discrete Cosine Transform (DCT) and other uniform filter banks satisfy the criterion of controlling the frequency location of the quantization distortion but do not provide a good model of the human visual system's structure. This presents a difficulty in creating masking models since there is a mismatch between the underlying structure of the model and the structure of the transform. However, it is worth studying how visual models can be utilized in a DCT framework, since it is the current framework for still image and video compression standards. The *Cortex Transform* [13] was introduced as a way to produce a transform that corresponds to the known structure of the human eye; a corresponding coder based on this transform is described in [15]. The main stumbling block in using this approach is that the Cortex Transform is not a maximally decimated filter bank. This results in an expansion in the amount of data that needs to be encoded. At present, visual masking models have not been able to provide enough coding gain to overcome this disadvantage.

JPEG is the current international standard for color still image compression [16]. The most basic version of this coder, referred to as the Baseline Sequential Codec, consists of decomposing the original image into non-overlapping blocks, applying a DCT to each block, quantizing the transform coefficients and entropy coding the quantized data. It is very desirable to see how we can take advantage of visual models within this framework, although it has already been mentioned that block-based DCTs are not ideal in terms of mimicking the human visual system's structure.

For JPEG compression, the original image is decomposed into non-overlapping 8 x 8 blocks and the DCT is performed independently for every block of data. Due to the block structure in the

decomposition, we refer to the original image as $x_{i,j,b}$ where *i,j* denotes the location in block *b* and $X_{u,v,b}$ denotes the DCT coefficient for the basis function associated with position *u,v* in block *b*. The quantizer step size that is used for each coefficient is given by a quantization table which is specified as an input to the encoder. Since JPEG allows the user to specify a quantization table for each image, it should be possible to derive a "perceptually optimal" one. In fact, in [17] a set of formulas for determining the perceptually optimal quantization matrices for both luminance and chroma given the image size, monitor white point, and viewing distance is presented. These formulas were derived by running a large set of subjective experiments to determine the detectability of each DCT basis function.

From these experiments, a perceptual model for predicting the detection thresholds based only on the viewing conditions and global properties of the visual system was derived. This model takes into account frequency sensitivity in determining the optimum quantization matrix but does not take into account the image dependent components of luminance sensitivity and contrast masking. This has been addressed in [18], where this approach has been extended to determine an image dependent quantization table that incorporates not only the global conditions, but also accounts for local luminance and contrast masking. An iterative approach is presented to determine an image dependent quantization table that provides a specified level of visual distortion. Since JPEG allows for only one quantization matrix for all the image blocks, it is difficult to take full advantage of the local properties as given by the model in [18].

The work in [19] introduces additional local quantizer control by using the previous model to drive a prequantizer which zeros out all DCT coefficients below the locally derived JND threshold. Such an approach is compliant with the JPEG bitstream specification while allowing some amount of local control as given by the visual model. The perceptual coder with the optimized quantization matrix and adaptive prequantizer is shown to yield significant improvement over the standard implementation of JPEG. Other perceptually based algorithms that have been proposed for image compression include [20] which is based on a subband decomposition and quantization step sizes derived from frequency and luminance sensitivity, and contrast masking. A perceptually based EZW algorithm is proposed in [21]. An overview of using visual models for signal compression is presented in [22].

A different image compression model has been developed in [2,23]. This model uses both frequency sensitivity, and spatial masking based on an image's edges. The spatial masking is a modified version of the spatial masking model presented in [24]. The main principle of the spatial masking is that edges in an image are able to mask signals of much greater amplitude than regions of near-constant intensity. For a given image, a *tolerable-error level* value (TEL) may be formed for each pixel. This quantity is similar in concept to the JND value. The TEL provides the allowed magnitude that a pixel can change without the changes becoming visible. The development of the TEL is further described in [2,23]. A third model, which is based on contrast masking is described in [25,26].

The JPEG bitstream specification limits the amount of perceptual fine-tuning that can be incorporated into the coder [16]. This applies to other coding schemes too, where the information obtained from an image dependent visual mask makes the amount of overhead needed to transmit the side information prohibitively large. Specifically, if the visual mask provides the

amount of just noticeable distortion that can be tolerated at every location, we can adapt the quantization step to be optimum everywhere. In practice, the quantization step size is chosen based either on the most sensitive portion of the image for transparent quality, with the data rate depending on the perceptual complexity of the image, or some average distortion over a specified region with a fixed data rate and variable quality. Many times, a very good visual model will reduce to a simple frequency weighing due to the limitations of the coder design. However, for the watermarking application we are not limited by the amount of bits needed to transmit the perceptual mask. Some of the models introduced in this section for image compression will be utilized later in developing very effective watermarking schemes.

3   PERCEPTUAL WATERMARKING FOR STILL IMAGES

3.1   Motivation

There are two basic modalities for image watermark encoding: spatial-domain techniques (*spatial watermarks*) and spatial frequency-domain techniques (*spectral watermarks*). This section first describes several spatial watermarking algorithms that rely on some type of perceptual knowledge in the encoder. Many of the spatial watermarking techniques provide simple and effective schemes for embedding an invisible watermark into the original image but are not robust to common image alterations. In the simplest sense, these are the first perceptually based watermarking techniques which rely on a scheme for watermark encoding which will produce resulting images of high quality but not necessarily robust to attacks. Another way to mark an image is to transform it into the frequency domain – Fourier, DCT, wavelet, etc. – before marking it. The mark is incorporated directly into the transform coefficients of an image. The inverse-transformed coefficients form the marked image. These types of algorithms are often called *spectral watermarks*, and commonly use frequency sensitivity of the human visual system to ensure that the watermark is invisible. Many of these techniques are *private watermarks*, which require the original image to verify the mark. Algorithms that do not require the original image for testing are called *public watermarks*. Many of these frequency-sensitivity based watermarking techniques, however, are not image-adaptive. Frequency weighting can be considered a property of the human visual system and viewing conditions, and does not necessarily have to adapt to individual image characteristics. The algorithms described at the end of this section are image-adaptive in that they use formal visual models to determine how to embed the watermark, and provide perceptual transparency as well as robustness to attack.

3.2   Perceptual Watermarking based on the Transparency Criterion

3.2.1   Checksum technique

An early watermarking algorithm, known as the checksum technique [27],is formed from the checksum value of the seven most significant bits of all pixels in an image. A checksum is the modulo-2 addition of a sequence of fixed-length binary words; it is a special type of hash function [6]. In this technique, one word is the concatenation of the seven most significant bits from eight consecutive pixels. Each pixel is involved in the checksum only once, and the final checksum length is fifty-six bits. The technique then randomly chooses the locations of the pixels that are to contain one bit of the checksum. The last bit of each chosen pixel is changed (if necessary) to equal the corresponding checksum bit. The pixel locations of the checksum

together with the checksum value itself form the watermark, *W*, which must be kept secret. To verify a test image *Z* the checksum of *Z* is computed, and compared to the checksum value in *W*. Any discrepancy means that the image *Z* is not an exact copy of *Y*.

3.2.2   Basic "spread spectrum" approach

The watermark in [28] is based on spread spectrum communications, which delivers narrowband data through a noisy channel, by modulating the each data symbol with a wideband (but very low amplitude) signal [29]. The data here is a single bit – a yes or no decision on whether the given watermark is present. The channel is the image data itself, and the wideband signal is the watermark. A linear feedback shift register with *n* stages can form pseudo-random binary sequences with periods as large as $2^n - 1$. M-sequences achieve this maximum period, and have very desirable autocorrelation and randomness properties [6]. Two types of sequences that may be formed from an m-sequence are *unipolar* and *bipolar*. The elements of a bipolar sequence are {-1,1} and the elements of a unipolar sequence are {0,1}. Let *X* be a grayscale 512 x 512 image, and *w* a bipolar extended m-sequence row of length 512. An extended m-sequence has properties similar to regular m-sequences [30]. One simply appends a zero to the end of the longest run of zeros in an m-sequence to create the extended version. *W* consists of 512 circularly shifted copies of *w* with random phase. *W* is then arithmetically added to *X* to form the watermarked image, *Y*.

$$Y = X + W \tag{1}$$

To verify a possibly forged image row *z* relative to the original row *y*, the spatial crosscorrelation function between *z* and *w* is obtained.

$$R_{zw}(\alpha) = \sum_j [z(j) - E[z]] \, w(j - \alpha) \tag{2}$$

E[z] is the average pixel value of row *z*. The presence of a peak in R_{zw} for a given offset α is determined. If there is no peak, *z* is not authentic.

This watermark is robust to small amounts of noise introduced in the image (i.e. the peak of R_{zw} is still noticeable). It can also accommodate multiple watermarks. However, the mark is not robust to random bit flipping in the lower two bit planes, or more sophisticated image processing attacks.

The watermarking scheme proposed in [31,32] is known as the *Variable-W Two-Dimensional Watermark* (*VW2D*). The authors reshape an m-sequence into two-dimensional watermark blocks, which are added and detected on a block-by-block basis. Neither *VW2D* nor [28] require the original image for watermark detection. VW2D, however, can detect local alterations in a received image on a block-wise basis. Many of the other published techniques address only the binary question of whether a given digital signature is present in a particular received image.

A patented technique embeds a different watermark for each single bit of information that an owner wants to encode in the image [33]. For instance, embedding a 16-bit user identification string would require 16 independent watermarks. The number of bits that can be embedded in this way is limited by an image's capacity with respect to this technique. Research is underway to quantitatively determine an image's capacity for certain watermarking algorithms [34].

3.2.3 Other spatial domain watermarks

Many other spatial domain watermarks exist. One algorithm partitions the original image into two sets, *A* and *B*, according to a pseudo-random partition *S* [35]. The luminance of pixels in set *A* is increased by an integer *k*. *k* is kept small enough to maintain the imperceptibility of the alterations. Verification is performed with *S* and *k*; the original image is not required. A similar method (called the patchwork method) chooses a set of pairs of pixels, and increases the difference between the value of each pixel in the pair [36]. A second method presented in [36] is called texture block coding. A watermark consisting of a textured patch is embedded into an area of the image with the same texture. This is a good example of using some common sense rules to determine where signal alterations will be least noticeable, which results in good visual quality for the marked image. This scheme, however, requires that the image contain relatively large areas of texture; the technique is also vulnerable to low-pass filtering. In other words, the transparency requirement comes at the expense of robustness.

Some invisible spatial watermarks exploit the fact that humans detect changes in different colors unequally. One such algorithm embeds sinusoidal patterns exclusively in the yellow-blue plane of an image [37]. Another performs amplitude modulation only in the blue plane [38]. Several invisible spatial watermarks have been created, which reside in all three color planes of an image [39,40]. A spatial watermark that embeds an image dependent graph in the image is presented in [41].

Another method selectively filters a certain subset of pixels in an image (called signature pixels) to embed a watermark [42]. The locations of signature pixels are determined by the watermark, *W*; to encode *W*, the pixel is replaced with a filtered version of itself. Detection is performed by first filtering the image, then identifying the pre-filtered pixels. One can then reconstruct *W*. The image is verified either by inspection of the recovered *W*, or by computing the normalized correlation between the recovered and original *W*. Another technique that only marks a subset of an image – in this case a contiguous band of rows of random start position and user-determined length – is described in [43].

A watermarking algorithm has been developed specifically for halftone images [44]. This algorithm exploits the fact that many different halftone patterns will produce a perceptually similar gray field in an image. The halftone pattern of an image is modified to incorporate a watermark, and the watermark is verified by inspection as follows: a transparent sheet (called the screen) with a certain halftone pattern is overlaid on a printed version of the watermarked image. Upon sliding the testing sheet into alignment with the printed image, a visible watermark appears. The mark is invisible in the printed image itself.

One perceptual spatial algorithm [45] segments the image into blocks, much like *VW2D*. The amplitude of the mark (which originates from a pseudorandom sequence that uniquely identifies the owner) is adapted to the corresponding image block. However, the amplitude of mark is the same for all pixels in the block.

3.3 Perceptual Watermarking based on Frequency Sensitivity

3.3.1 A spread spectrum watermark embedded in the DCT domain

A frequency domain method for digital watermarking of images proposed in [46] is also based on the idea of spread spectrum communications. The technique is motivated by both perceptual transparency and watermark robustness. The published results show that the technique is very effective both in terms of transparency, robustness to signal processing, and attempts to remove the watermark. The types of image distortions to which technique is robust include cropping, very low data rate JPEG compression, printing and rescanning, as well as collusion with several independently watermarked images. One of the significant contributions in this work is the realization that the watermark should be inserted in the *perceptually significant* portion of the image in order to be robust.

The watermark W is a sequence of normally distributed, zero-mean unit-variance random numbers. A DCT is performed on the entire image and W is inserted in a predetermined range of low frequency components minus the DC component as follows. Let X be the original image, Y be the watermarked image, and X_D and Y_D be the DCT coefficients of X and Y respectively. The coefficients can be ordered according to the zigzag pattern used in JPEG. Let $X_D(i)$ and $Y_D(i)$ be the ith DCT coefficient in X_D and Y_D respectively. $W(i)$ is the ith element in the watermark sequence; a is a scale factor which prevents unreasonable values for $Y_D(i)$. The marking is then performed:

$$Y_D(i) = X_D(i)(1 + aW)$$
(3)

Alternate equations that incorporate W into X_D are listed in [46]. Inversely transforming Y_D to form Y completes the marking procedure. Figure 2 shows an original image (upper left), marked with $a = 0.1$, 0.5 and 1.0 respectively. The authors propose an empirically derived value of 0.1 for a.

The first step of the verification procedure is to obtain a copy of W from a possibly forged image Z. Z_D is the vector of Z's DCT coefficients. W^* is the extracted version of W.

$$W^*(i) = \frac{1}{a}\left[\frac{Z_D(i)}{X_D(i)} - 1\right]$$
(4)

A measure of similarity between W^* and W is computed as follows:

$$S(W, W^*) = \frac{W \cdot W^*}{\sqrt{W^* \cdot W^*}}$$
(5)

If an image has not been watermarked with W, S is distributed as a zero mean random variable. If W^* differs only slightly from W (i.e. W is indeed present in Z, although slightly altered), then $E[S] \gg 0$. A hypothesis test on S determines if W is present in the image. This technique accommodates multiple watermarks, and withstands a much wider range of attacks than the transparency-only based spatial techniques.

This algorithm is one of the earliest attempts at providing some image adaptability in the watermark embedding scheme. This is due to the fact that the watermark strength depends on the

intensity value of the DCT coefficients of the original image. In this way, the watermark signal can be quite strong in the DCT values with large intensity values, and is attenuated in the areas with small DCT values. This provides a watermark signal that is quite robust and for most images, transparent. However, because the DCT transform in this scheme is based on the whole image rather than the usual block-based approach commonly found in image and video compression schemes, the transform does not allow for any local spatial control of the watermark insertion process. In other words, the addition of a watermark value to one DCT coefficient affects the entire image; there is no mechanism for local spatial control in this particular framework. This scheme may benefit from a perceptual model that determines the optimal weights for the DCT coefficients, but the framework needs to be modified in order to get finer control of watermark adaptability to image characteristics and the human visual system.

Figure 2. Example of DCT-spread spectrum technique in [46];
from left to right, top to bottom: original, $a = 0.1$, $a = 0.5$, $a = 1.0$

Another global method also modulates DCT coefficients, but uses a one-dimensional bipolar binary sequence for W [47]. The DCT of the original image is first obtained. The marking procedure consists of sorting the DCT coefficients according to their absolute magnitude. The

owner then defines a percentage of total energy, P, and identifies the largest n coefficients that make up P percent of the total energy. The watermark sequence is then added to all the AC coefficients in this list. For all i such that $X_D(i)$ is one of the selected coefficients,

$$Y_D(i) = X_D(i) + W(i) \tag{6}$$

A larger P increases the number of elements of W that can be embedded in X, but increases the chance that W will be perceptible. W and the list of selected coefficients must be kept secret. The verification procedure first extracts W^* from the marked coefficients in Z_D:

$$W^*(i) = Z_D(i) - X_D(i) \tag{7}$$

A procedure similar to [46] can then verify W^*. Note that [46] and [47] both require X to extract the watermark.

3.3.2   A linear combination of marked and unmarked images

The method in [48] is similar to [46]. The DCT of an entire original image X is computed, and the coefficients are ordered in the zigzag fashion of JPEG to form the coefficient vector X_D. To decrease the chance of the watermark being perceptible, the first L coefficients are not marked. W is a length M pseudorandom sequence of numbers, which is added to DCT coefficients $X_D(L+1)$ to $X_D(L+M)$. L and M are user-defined.

$$A_D(k) = \begin{cases} X_D(k) + \alpha \left| X_D(k) \right| W(k), & k = L+1, L+2 \ldots L+M \\ X_D(k), & otherwise \end{cases} \tag{8}$$

A_D forms the vector of modified coefficients, and α is a user-defined scaling factor. The inverse transform of A_D forms an intermediate image, A. Pixels in the marked image, Y, are linear combinations of the pixels in X and A. For each pixel $y_{i,j}$ in Y:

$$y_{i,j} = (1 - \beta_{i,j})x_{i,j} + \beta_{i,j}a_{i,j} \quad \beta \in [0,1] \tag{9}$$

The pixel $a_{i,j}$ is the pixel at the ith row and j^{th} column in image A. $\beta_{i,j}$ is the normalized variance of the pixels in a 9 x 9 neighborhood of pixel $x_{i,j}$. This variance has been normalized with respect to the maximum variance of all such 9 x 9 regions. By using β, the scaling factor α can be higher than if A were the final watermarked image. The authors suggest using $\alpha = 0.2$. Testing is similar to [46].

3.3.3   Other transform-based approaches

An early DCT-based technique is presented in [49,50]. The image is segmented into 8x8 non-overlapping blocks; each block is transformed into the frequency domain using the DCT. This is the same building block that is used in JPEG image compression. A pseudorandom subset of the blocks is chosen and a triplet of midrange frequencies is slightly altered to encode a binary sequence. This seems to be a reasonable approach for adding some sort of perceptual criterion. Watermarks inserted into the high frequencies are most vulnerable to attack whereas the low frequency components are perceptually significant and very sensitive to alterations; such alterations may make the watermark visible. This scheme should provide reasonable results on average although a more image dependent scheme could provide better quality and robustness. The original image is not required to extract the sequence.

The basic ideas introduced in [49] are further extended in [51] by introducing the watermark encoding in the actual quantization process of the mid-frequency coefficients. The result is two schemes that also do not require the original image for watermark decoding. The first scheme embeds a linear constraint among the selected DCT coefficients; the second defines a circular detection region in the DCT domain similar in concept to vector quantization (VQ) [52]. A different improvement to [49] is presented in [53,54]. The original image is segmented into 8 x 8 blocks. Image blocks that contain either sharp edges, or have little texture are not marked. In these blocks the watermark in [49] would more easily be perceived. A similar technique that classifies blocks of the image according to their energy content is described in [55]; the amount of energy in the block determines in part the amplitude of the mark to be embedded in that block.

Visual inspection is performed on a technique which embeds a non-random image (i.e. a Kanji character) into an image [56,57]; again only the middle frequency DCT coefficients are modified. Another method [58] embeds the watermark in the phase information of the Discrete Fourier Transform (DFT) of an image. A second DFT-based technique uses properties of the DFT to create a watermark resistant to geometric attacks (rescaling, translation and rotation) [59]. One such property is the rotation-invariance of the DFT amplitude [60]. The second method presented in [45] sets certain DCT coefficients to zero. This is done when the corresponding watermark bit equals 1; only higher frequency coefficients are adjusted. A different technique assigns an integer to each 8 x 8 DCT block; adding this integer to all non-zero DCT coefficients in the block marks the block [61].

A wavelet-based version of [46] is described in [62], which embeds the watermark according to the following equation:

$$Y_{l,f}(u,v) = X_{l,f}(u,v) + \alpha\big[X_{l,f}(u,v)\big]^2 W(u,v) \qquad (10)$$

where $X_{l,f}(u,v)$ refers to the wavelet coefficient at position (u,v) in resolution level l and frequency orientation f, and α is a user-determined constant that is maximized under the transparency constraint.

3.4 Perceptual Watermarking based on Image-Adaptability

3.4.1 Main principles: transparency, robustness and capacity

We begin by reviewing some of the requirements that are necessary to provide a useful and effective robust watermarking scheme. We will briefly discuss the requirements as an introduction to the use of perceptual information in meeting these requirements. Three requirements (or features) for such watermarking schemes are *transparency*, *robustness* and *capacity*. Other requirements which may be important for watermarking, depending on the specific application, include: being able to detect the watermark without the original image, watermark encoder/decoder complexity, the ability to encode a watermark directly into a compressed bitstream, tamper detection, localization of changes to an image, etc. Transparency refers to the perceptual quality of the data being protected. For the case of image data, the watermark should be invisible over all image types as well as local image characteristics. Such a requirement is most challenging for images composed of large smooth areas where it is very

difficult to introduce modifications to the pixel intensities without affecting the overall image quality.

The digital watermark should also be robust to alterations to the image, either friendly or unfriendly, which could alter or remove the signature. Ideally, the amount of signal distortion necessary to remove the watermark should degrade the desired image quality to the point of becoming commercially valueless. Possible image alterations include intentional transformations of the image data as well as illegal attempts to remove the mark, or transform the watermark into another valid watermark. Typical image transformations include compression, in particular JPEG for still images and MPEG for video, resampling, requantization, image enhancements, cropping, and halftoning. Capacity may also be a critical feature for applications where the watermark identifies the buyer or end-user. Capacity refers to the ability to detect watermarks with a low probability of error as the number of watermarks in a single image increases. The watermarking technique should provide a way to insert the maximum number of distinguishable watermarks for the given technique. To best meet these three requirements, we would like the watermark to adapt to the local image characteristics as well as viewing conditions in order to provide the strongest signature without compromising the image quality.

3.4.2   Differences between image-adaptive and transparency-based watermarks

The above requirements of transparency, robustness and capacity introduce a challenging problem from the signal processing perspective. The most straightforward way to introduce a transparent watermark results in a watermark that is very vulnerable to attack. For example, placing a watermark in the least significant bits or in the high frequency components results in very good image quality but can be destroyed with simple quantization or lowpass filtering; in other words, from a signal processing viewpoint, the requirements of transparency and robustness conflict with each other. A similar conflict exists from a perceptually based viewpoint. In order to provide transparency it makes sense to consider *perceptually insignificant* components for watermark insertion. However, to provide robustness it makes sense to insert a watermark in the *perceptually significant* components for watermark insertion.

3.4.3   Use of formal perceptual models

We would like to embed an imperceptible digital signature into an original image that is difficult to remove without destroying the perceived original image quality. For applications where we would like to label the image with a watermark that uniquely identifies the end-user, we may also wish to provide the maximum number of unambiguous watermarks. The use of either formal visual models or common sense rules based on some knowledge of the human visual system would be beneficial in developing watermark encoders that provide transparent quality. This is a common theme in many of the published techniques. It also turns out, as will be described in detail later in the paper, perceptual information could also be used to develop a robust watermarking scheme. By using formal visual models, we describe several techniques that can adapt the watermark to each image. In theory, a good visual model should provide the maximum strength, maximum length watermark sequence that can be inserted without introducing visual distortions. This is also the best we can hope to achieve in terms of capacity, subject to the invisibility and robustness constraints. The techniques described below use formal visual models,

some of which were described in Section 2 and developed originally for image compression applications.

3.5 Overview of Image Adaptive Watermarking Techniques

3.5.1 Image adaptive-DCT and image adaptive-wavelet algorithms

The two techniques described here, the image-adaptive DCT (IA-DCT) approach [63,64] as well as the image-adaptive wavelet (IA-W) approach [65] have been motivated by the excellent results presented in the spread spectrum technique of [46]. The authors introduced the use of formal visual models, into two watermarking frameworks. The goal is to investigate if a more image-adaptive scheme has any advantages over the very good results that have already been achieved with much simpler schemes that only use image adaptability on a global level.

The frequency decomposition for the image-adaptive DCT algorithm is based on an 8 x 8 DCT framework. Unlike the decomposition in the spread spectrum approach [46], the block-based approach provides local control that allows for incorporating local visual masking effects. A benefit of such a scheme (and selected other block-based DCT schemes) is that if the images are stored as compressed JPEG bit streams, the watermarks can be inserted directly to the partially decompressed bit stream [54,64].

It is important to note as described by the authors, that for the watermarking problem, all the local information extracted from the visual models can be utilized in determining the watermark encoding algorithm. The local information is stored in what is called a *just-noticeable difference* matrix (JND). The values in the JND matrix are based on the frequency domain representation of the image; they are the thresholds beyond which any changes to the respective coefficient will most likely be visible. This is in contrast to compression applications, where local adaptation is minimal due to the tremendous amount of overhead required to send the JND information. In the applications addressed here, the original image is available at the decoder and the JND threshold values can be obtained directly from this image. Actually, the JND thresholds can be estimated from the received watermarked image fairly accurately; this means that this technique can be applied to applications where the original image is not available for watermark detection.

The image-adaptive DCT-based approach uses one of the visual models described in Section 2 for image compression in a JPEG framework [18]. Recall that the JND thresholds derived from the visual model consist of an image independent part based on frequency sensitivity, and an image dependent part based on luminance sensitivity and contrast masking. These three components of the visual model have been described in detail in Section 2. They have been derived in the context of image compression to determine the maximum amount of quantization noise that can be tolerated at every image location without affecting the visual quality of the image (under the specific viewing conditions used in the model). In the context of image watermarking, the JND thresholds can be used to determine the maximum amount of watermark signal that can be tolerated at every image location without affecting the visual quality of the image. *W* consists of a sequence of real numbers generated from a Gaussian distribution with zero mean and unit variance as proposed in the spread spectrum technique of [46].

We present two watermarking schemes: the IA-DCT and IA-W scheme. The watermark encoder for both the IA-DCT and IA-W schemes can be described in general as

$$X^*_{u,v} = \begin{cases} X_{u,v} + J_{u,v}w_{u,v}, & X_{u,v} > J_{u,v} \\ X_{u,v}, & otherwise \end{cases} \quad (11)$$

where $X_{u,v}$ refers to the frequency coefficients of the original image samples $x_{i,j}$, $X^*_{u,v}$ refers to the watermarked image coefficients, $w_{u,v}$ is the sequence of watermark values, and $J_{u,v}$ is the computed *just noticeable difference* calculated for each frequency coefficient. A block diagram of the general image-adaptive perceptual watermarking scheme is illustrated in Figure 3. We describe two techniques: one where the frequency decomposition consists of block-based DCTs, and one where the frequency decomposition consists of a wavelet transform.
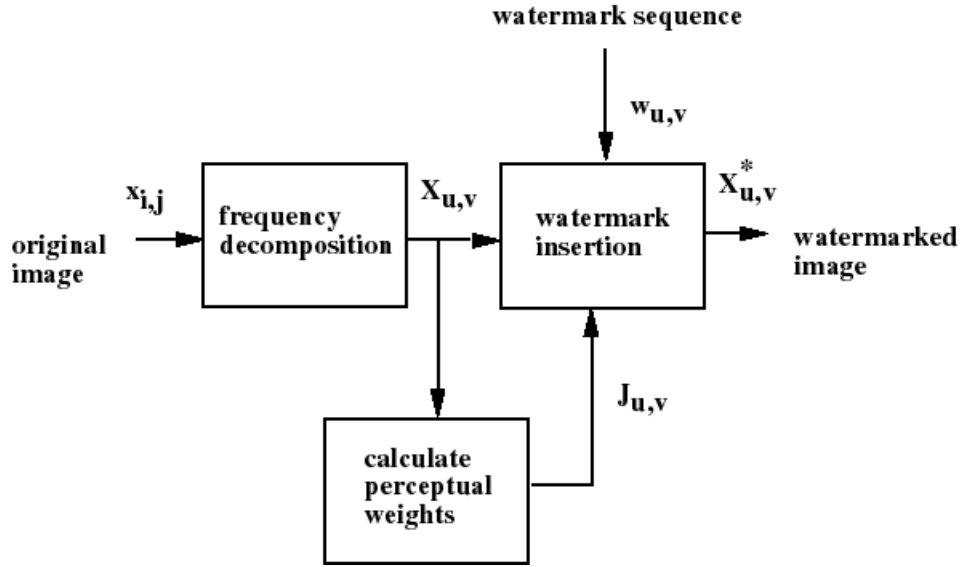

Figure 3. General block diagram of the IA watermarking schemes.

The watermark encoder for the IA-DCT scheme is described as

$$X^*_{u,v,b} = \begin{cases} X_{u,v,b} + J_{u,v,b}w_{u,v,b}, & X_{u,v,b} > J_{u,v,b} \\ X_{u,v,b}, & otherwise \end{cases} \quad (12)$$

where as before, $X_{u,v,b}$ refers to the DCT coefficient at position *(u,v)* of block *b*, $X^*_{u,v,b}$ is the watermarked DCT coefficients, $w_{u,v,b}$ is the sequence of watermark values and $J_{u,v,b}$ is the computed *just noticeable difference* calculated from the model in [18] as described in Section 2. At times we do have *a priori* knowledge about some of the image transformations that will be applied to the watermarked image and it would be beneficial to take advantage of this knowledge in the watermark encoding process. For instance, if we know that the image will be lowpass filtered, it is best to avoid placing the watermark sequence in the high frequency components. In general, however, watermark insertion is not limited only to perceptually significant parts of the image and the authors assume no *a priori* knowledge about friendly or unfriendly image alterations. The goal in this approach is to take advantage of the full capacity of the image in

18

order to place a maximum strength watermark sequence that will be very difficult to alter or remove.

Note that since the watermark is generated from a normal distribution, watermark insertion will occasionally result in values that exceed the JND. Informal studies show that exceeding the JND occasionally does not result in any visibly objectionable results. This might signify that there are other masking effects that could be incorporated into the visual models that we are not currently taking advantage of. Currently, the watermark is only inserted into the luminance component of the image.

For the IA-W scheme [65], frequency sensitivity thresholds are determined for a hierarchical decomposition using the 9-7 biorthogonal filters in [66]. Due to the hierarchical decomposition, this approach has the advantage of consisting of watermark components that have varying spatial support. This provides the benefits of both a spatially local watermark and a spatially global watermark. The watermark component with local spatial support is suited for local visual masking effects and is robust to signal processing such as cropping. The watermark component with global spatial support is robust to operations such as lowpass filtering. Due to the hierarchical nature of such an approach, this scheme is more robust to certain types of distortions than the DCT-based framework.

The wavelet framework consists of a four level decomposition as illustrated in Figure 4. Here, the upper left-hand corner corresponds to the lowest frequency band. The visual model used here is much simpler than the one used in the DCT-based scheme. A wavelet-based JND value $J_{l,f}$ is determined for each frequency band based on typical viewing conditions. Here l denotes the resolution level where $l = 1,2,3,4$ and f denotes the frequency orientation where $f = 1,2,3$. Referring to Figure 4, the frequency locations 1 and 2 refer to low horizontal/high vertical frequency components and low vertical/high horizontal frequency components respectively. Frequency location 3 refers to high horizontal/high vertical frequency components. The details of the experiments and resulting weights can be found in [67].

Adding image dependent components as in the DCT-based approach could further refine this model. However, even this simple visual model yields very good results and the hierarchical framework provides a robust watermark as well as finer control of watermark insertion than can be obtained using a block-based scheme. Results comparing the wavelet-based scheme to the DCT-based scheme are described in [65].

The watermark encoder for IA-W is described by

$$X^*_{u,v,l,f} = \begin{cases} X_{u,v,l,f} + J_{l,f}\, w_{u,v,l,f}, & X_{u,v,l,f} > J_{l,f} \\ X_{u,v,l,f}, & otherwise \end{cases} \tag{13}$$

where $X_{u,v,l,f}$ refers to the wavelet coefficient at position (u,v) in resolution level l and frequency orientation f, $X^*_{u,v,l,f}$ refers to the watermarked wavelet coefficient, $w_{u,v,l,f}$ is the watermark sequence and $J_{l,f}$ corresponds to the computed frequency weight at level l and

frequency orientation *f* for the 9-7 biorthogonal filters. As for IA-DCT, the watermark is inserted only in the luminance component of the image.
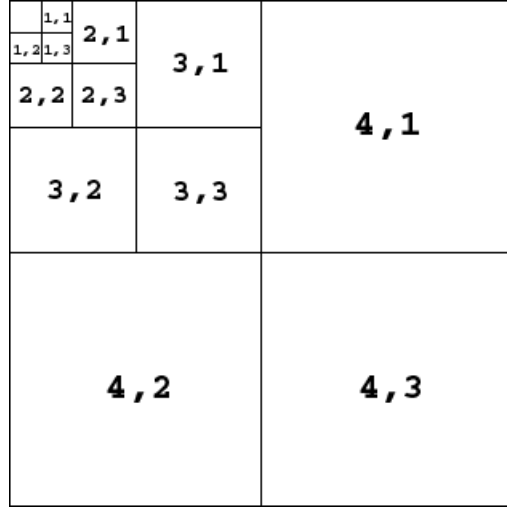

Figure 4. Four level wavelet decomposition

Watermark detection for the spread spectrum approach as well as the IA-DCT and IA-W schemes is based on classical detection theory [68]. The received image is subtracted from the original image and the correlation between the signal difference and a specific watermark sequence is determined. The correlation value is compared to a threshold to determine whether the received image contains the watermark in question. The normalized correlation detection scheme for the IA-DCT scheme can be expressed in vector space as:

$$w^*_{s,u,v,b} = X_{u,v,b} - \hat{X}^*_{u,v,b} \tag{14}$$

$$w^*_{u,v,b} = \frac{w^*_{s,u,v,b}}{J_{u,v,b}} \tag{15}$$

$$\rho_{ww^*} = \frac{w^* \cdot w}{w^* \cdot w^*} \tag{16}$$

where $w \cdot w^*$ denotes the dot product, $w^*_{s,u,v,b}$ denotes the possibly distorted watermark scaled by the JND thresholds $J_{u,v,b}$, $w^*_{u,v,b}$ denotes the received watermark, and ρ_{ww^*} is the normalized correlation coefficient between the two signals w and w^*. If w is practically identical to w^* and normally distributed, the correlation coefficient approaches 1. If w and w^* are independent, ρ_{ww^*} is also normally distributed. Therefore, the probability of ρ_{ww^*} exceeding a certain threshold can be directly calculated from the normal distribution. Comparing the correlation coefficient to a threshold value T_ρ performs the watermark detection. This threshold can be modified according to the tradeoff between the desired probability of detection, P_D, and the probability of false identification (false alarm), P_F. The final step for watermark detection is

$$\rho_{ww^*} > T_\rho \quad \textit{watermark w detected}$$
$$\rho_{ww^*} \leq T_\rho \quad \textit{watermark w not detected} \tag{17}$$

Any prior knowledge about the image transformations should be incorporated either at the encoder or decoder. For instance, if it is known that the image is to be lowpass filtered in some way, the high frequency components should be avoided for watermark encoding. At the decoder, the potential watermark sequence should be "whitened" before the correlation step in order to achieve better detection results. The work presented by [46] offers several techniques to estimate the channel degradations given the received, possibly watermarked image and the original image at the receiver.

Similarly, the watermark decoder for the wavelet scheme is also based on a correlation receiver. What is different in the IA-W decoder is that the correlation is computed separately for each resolution level as well as each frequency bin as labeled in Figure 4.

$$w^*_{s,u,v,l,f} = X_{u,v,l,f} - \hat{X}^*_{u,v,l,f} \tag{18}$$

$$w^*_{u,v,l,f} = \frac{w^*_{s,u,v,l,f}}{J_{l,f}} \tag{19}$$

$$\rho_{ww^*}(l,f) = \frac{w^*_{l,f} \cdot w_{l,f}}{w^*_{l,f} \cdot w^*_{l,f}}, \quad l = 1,2,3,4 \; and \; f = 1,2,3 \tag{20}$$

In this case the normalized correlation is calculated separately for each subband (l,f). The average for each resolution level l is computed as

$$\rho_{ww^*}(l) = \frac{1}{N_f} \sum_{f=1}^{N_f} \rho_{ww^*}(l,f), \quad l = 1,2,3,4 \tag{21}$$

where N_f is the number of frequency orientations. In this case $N_f = 3$. Evaluating the correlations separately at each resolution can be used to our advantage in the detection process. For instance, cropping the image will impact the watermark values in the lower layers more than in the higher layers. This is due to the fact that the bands in higher layers (and the corresponding watermark sequence) correspond to a smaller spatial support. Likewise, any type of lowpass filtering operation will affect the higher layer watermark coefficients more than the lower layer coefficients. In this case the layers with low correlation values would be discarded from the computation of the average. Similarly, the average correlation value over a certain frequency orientation may be computed.

$$\rho_{ww^*}(f) = \frac{1}{N_l} \sum_{l=1}^{N_l} \rho_{ww^*}(l,f), \quad f = 1,2,3 \tag{22}$$

N_l is the number of frequency orientations; in this case $N_l = 4$. By evaluating the correlations separately for each frequency location, one can take advantage of any strong structure that is associated with the original image where the watermark sequence was much stronger than in

other parts of the image. The maximum correlation value over all the possible levels as well as frequency locations may also form the test statistic.

$$\rho_{ww^*}^* = \max_{l,f}\left\{\rho_{ww^*}(l), \rho_{ww^*}(f)\right\} \tag{23}$$

3.5.1.1 Experimental results for IA-DCT and IA-W

Figure 5 and Figure 6 show images marked with IA-DCT (first rows) and IA-W (third rows). The corresponding watermark is directly underneath the image. The watermarks in certain images benefit greatly from the image-adaptive nature of the marking procedure. The images in Figure 5 are fairly uniform (in texture, brightness, etc.), and cannot take full advantage of the image-adaptive nature of the watermark. Figure 6 illustrates images that contain non-uniform properties where image-adaptability is particularly useful. This usefulness is shown by the large areas of very-low watermark amplitude, together with areas of very large amplitude. In Figure 5 the watermarks are much less distinct. Note that if a non image-adaptive marking method were used for the images of Figure 6, the maximum amplitude of the mark would have to be much lower than that of the IA algorithms in order to keep the mark imperceptible in the regions of little texture.

3.5.1.2 Robustness to Image Processing Operations

Here we discuss in more detail how the perceptually based watermarking algorithms that have been designed to be robust to attack perform under some typical image processing operations. More detailed robustness experiments and comparisons for the IA-DCT and IA-W approaches described here can be found in [65]. A detailed set of experiments for the spread spectrum approach can be found in [46]. It has been noted earlier that the requirements of transparency and robustness to attack conflict with each other. By designing watermarking algorithms that take advantage of properties of the human visual system, the goal is to produce extremely robust watermarking schemes while being able to guarantee transparency under particular viewing conditions.

3.5.1.3 JPEG compression

In the digital domain, it is very likely that the original source material may be compressed for more efficient storage or transmission. It is therefore important to examine whether the watermarking schemes can survive JPEG compression, since this is the current international standard for still image compression. The lowest data rate achievable for an image without producing any visible distortions is highly dependent on the original source material, and can range typically from 2 bits per pixel (bpp) for a highly complex scene to 0.2 bpp for a simple scene. An average rate of approximately 1 bpp produces reasonable quality for a typical image.

Figure 5. IA-DCT watermarked images (top row), IA-W watermarked images (third row) with corresponding IA-DCT watermarks (second row) and IA-W watermarks (fourth row)
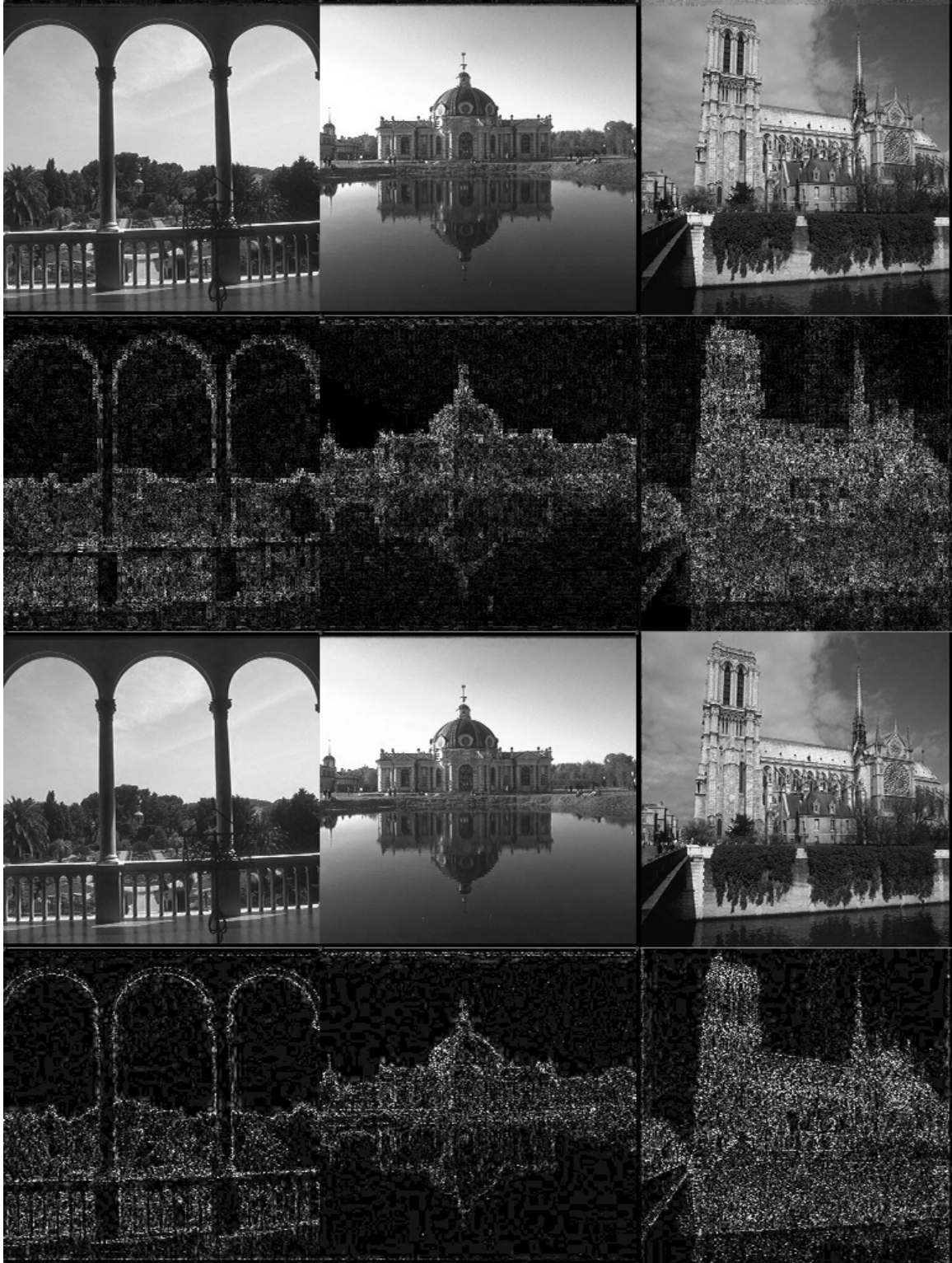
Figure 6. IA-DCT watermarked images (top row), IA-W watermarked images (third row) with corresponding IA-DCT watermarks (second row) and IA-W watermarks (fourth row)

Figure 7 illustrates a watermarked image that has been JPEG compressed to different data rates. The uncompressed watermarked image appears on the upper left side. The image on the top right is the watermarked image compressed from an original 8 bpp to 0.6 bpp. The bottom left image is compressed to 0.4 bpp, and the bottom right image to 0.18 bpp. Table 1 shows the watermark detection values for the IA-DCT scheme at different data rates. The detection results are very good for JPEG compression for all but the poorest quality images. Note that it would be sufficient to be able to detect the watermark only for the data rates that produce acceptable image quality. For instance, the compressed image at 0.18 bpp has little commercial value due to the poor image quality as shown in Figure 7. A more detailed study of robustness to JPEG compression can be found in [65] where compression results are shown for a variety of images at different data rates comparing the spread spectrum approach with the IA-DCT and IA-W approaches. All three schemes are shown to be effective against JPEG compression with the IA-W approach yielding the overall best performance.

3.5.1.4 Rescaling

Another common image transformation that may occur to a watermarked image is rescaling. In this case, we examine whether the watermark can survive lowpass filtering followed by subsampling and interpolation. Figure 8 illustrates the undisturbed IA-DCT watermarked image on the upper left-hand side with a decimated version followed by interpolation shown in the other three examples. The image on the upper right-hand side has been decimated by two in each direction, the lower left hand image has been decimated by four in each direction and the lower right hand image has been decimated by eight in each direction. In addition, all of the images were previously converted from RGB space to YCbCr space. The watermark detection values for the three decimated examples are given in Table 2. Although the IA-DCT watermark survives the subsampling introduced here, the authors in [65] show that subpixel decimation or interpolation will result in poor watermark detection results for the IA-DCT scheme due to the coherent detection scheme.

3.5.1.5 Cropping

So far, we have tested examples where in some sense we are throwing away frequency components of the watermarked image signal. In one case we discard frequency information through compression, where the high frequency components are discarded in order to provide an efficient representation; in the second case we use lowpass filtering in order to avoid aliasing in the rescaling case. It is also interesting to examine the dual case, where we are discarding spatial (instead of spectral) information through cropping, which is a common editing operation. Figure 9 illustrates the undisturbed watermarked image on the top left followed by this image cropped to 1/4 of the original size on the upper right, 1/16 of the original size on the bottom left and 1/64 on the bottom right.

We assume that the original, unmarked image is available for watermark detection and we are able to register the cropped watermarked image with the original image. To register a cropped image, one replaces the missing image sections with the corresponding parts of the original (unmarked) image. Each cropping experiment results in perfect correlation for the IA-DCT scheme due to the local spatial control that this framework allows. This means that since we can register the original image with the cropped watermarked images, we only calculate the

25

correlation value only over the cropped area. If the frequency framework were based on a global transform, the watermark detection would be affected since any amount of cropping will affect every watermark value in the image. Cropping still affects the capacity of the image because the amount of cropping that is considered acceptable will determine the maximum length watermark that can be recovered from the image. For more details on these experiments as well as others, please refer to [65].

Figure 7. Uncompressed and JPEG compressed versions of the marked garden image
From left to right, top to bottom: uncompressed image, bpp = 0.6, 0.4 and 0.18

Table 1. Robustness of IA-DCT to JPEG compression for garden image

Bits per pixel	1.18	0.8	0.6	0.4	0.27	0.18
ρ_{ww^*}	0.98	0.61	0.47	0.29	0.17	0.08

Figure 8. Watermarked image and 3 scaled versions

Table 2. Robustness of IA-DCT to subsampling and interpolation

Scaling factor:	Unscaled	2	4	8
Test statistic:	1.0	0.9	0.75	0.66

Several techniques for removing the watermark have been discussed in the literature. The authors in [46] show that collusion is an effective way to destroy the watermark. One collusion method is to average several versions of the same image with different watermarks. The authors point out that averaging over ten such images would effectively wipe out all the watermarks and leave the original image unchanged.

Figure 9. Cropped versions of watermarked image

3.5.1.6 Capacity

Capacity refers to the ability to be able to distinguish multiple watermarks with low probability of error as the number of different watermarks in an image increases. In the IA watermarking procedures, a watermark element is ignored if the unmarked frequency coefficient is less than that coefficient's JND value. The length of an IA watermark for a particular image is therefore defined as the number of watermark elements actually incorporated into the image. A longer watermark sequence will provide more capacity than a shorter sequence. For this reason, using perceptual models to insert the longest watermark sequence adapted to each image under the constraint of transparency is an important goal. As an example, in Table 3 we show the lengths of the IA-DCT watermarks for the images in Figure 5 and Figure 6. Note how the length of the sequence varies for each image. This is in contrast to some of the other techniques such as the DCT-spread spectrum technique [46], where the authors suggest using a watermark sequence of length 1000 for all images. The lengths of the watermark sequences for the images in Figure 5 and Figure 6 are not surprising. The images with large smooth areas (such as Figure 5 (right) and Figure 6 (center)) carry shorter sequences than the images with large areas of texture (such as Figure 5 (center) and Figure 6 (right)). This example shows that for the IA schemes, capacity is highly dependent on the particular image characteristics. Other factors that affect the capacity

include the type and amount of degradation that the watermark should survive. An example is cropping, which directly affects the watermark capacity simply by reducing the size of the image available for testing. This issue as well as the collusion attack is discussed in further detail in [34].

Table 3. Length of IA-DCT watermark for images in Figure 5 and Figure 6

Figure	Left	Center	Right
Figure 5	22804	28504	15359
Figure 6	15160	14745	27278

3.5.1.7 Testing IA-DCT without the original image

A revision to IA-DCT as applied to JPEG images [64] is proposed in [69], and avoids the use of the original unmarked image in the verification procedure. In this technique, it is assumed that the original image has already been JPEG compressed. The marking procedure is similar to IA-DCT, except a different subset of DCT coefficients are marked. This subset is called the *feature vector*, denoted *{X_D}*. If a DCT coefficient, x_D , is larger than 1/2 of its corresponding quantization table value, *Q*, it is included in *{X_D}*:

$$x_D \in \{X_D\}, \quad x_D > \frac{Q}{2} \tag{24}$$

The watermark *W* is a sequence of *N(0,1)* random numbers that is added to *{X_D}*:

$$
\begin{aligned}
y_D &= x_D + w, \quad x_D \in \{X_D\} \\
y_D &= x_D, \quad x_D \in \{X_D\}
\end{aligned}
\tag{25}
$$

The IDCT of Y_D forms the marked image *Y*.

To verify the presence of *W* in a test image *Z*, the feature vector *{Z_D}* is obtained. A correlation measure, *c*, is found between *{Z_D}* and *W*. The mean and variance of the point-wise multiplication of *{Z_D}* and *W* are μ and σ.

$$\mu = \frac{\sum_{i=1}^{n} z_D(i)w(i)}{n} \tag{26}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} [z_D(i)w(i) - \mu]^2}{n-1} \tag{27}$$

$$c = \frac{\mu\sqrt{n}}{\sigma} \tag{28}$$

W is assumed to be uncorrelated with $\{Z_D\}$. Under certain conditions (large n, among others), c will be distributed roughly according to N(0,1) if W is not present in $\{Z_D\}$ [70]. If W is present then the expected value of c, μ_c, will be much higher. [Experimental results place μ_c at roughly 30 for a sample 512 x 512 image. The reader is referred to [69] for the details.] The same threshold test used in the IA-DCT method is performed on c to determine if the W under test is present in Z. If so, then Z belongs to the owner that corresponds to W.

3.5.2 Sub-band watermarking

This algorithm is described in [25], and is similar to the authors' earlier work in [71]. An original image, X, is first passed through a Gabor filter bank [72]; other perceptually tuned filter banks could be used. The energy in each band is obtained as shown in Figure 10. The center frequencies of each filter are indicated as f_k. E_k is the energy of the image present in the frequency band specified by filter k. The second step in marking X is to obtain a random two-dimensional array, and then low-pass filter it to form G. G is modulated at each center frequency, f_k to form G_k. The energy of G_k, labeled D_k, is obtained. The watermark can then be formed:

$$W = \sum_k \alpha_k G_k \tag{29}$$

$$\alpha_k = \begin{cases} 1, & E_k > D_k \\ 0, & E_k \le D_k \end{cases} \tag{30}$$

$$Y = X + W \tag{31}$$

The verification procedure first passes Y through the filter bank, and attempts to extract G from each subband that had (a modulated version of) G embedded in it originally. If, through the detection techniques mentioned previously, a sufficient number of subbands are determined to contain the watermark, the image belongs to the owner of W. This technique is robust to JPEG compression, and tolerates common attacks such as low-pass filtering. As in the IA-W technique, certain subbands may be excluded from consideration if a priori knowledge of certain attacks is available.

3.5.3 Spatial domain correction of an image adaptive watermark

The image watermarking technique in [73] first computes the DCT of an original image X on an 8 x 8 block-wise basis. Masking thresholds, m, are defined and computed for each block based on the DCT coefficients; these thresholds are similar in theory to the JND values used in IA-DCT. The watermark for an individual block, w, is a reshaped m-sequence. Different w are used for each block in X. The DCT of w is computed, and a scale factor s for the block is derived that is deigned to maximize the amplitude of w, while keeping each element in w below the corresponding value in m.

$$y_D = x_D + s \cdot w_D \tag{32}$$

The process is repeated for each block. To ensure that the addition of w is imperceptible, spatial domain correction is employed on the marked block y as follows. A tolerable error level (TEL) value (as described in Section 2) is computed for each pixel in x. If

$$|y - x| > TEL(x) \qquad (33)$$

for any pixel in y, s for the block is reduced and the corresponding block x is remarked with the lower scale factor. S is the entire array of masking blocks s. Figure 11 shows the block diagram for this method.
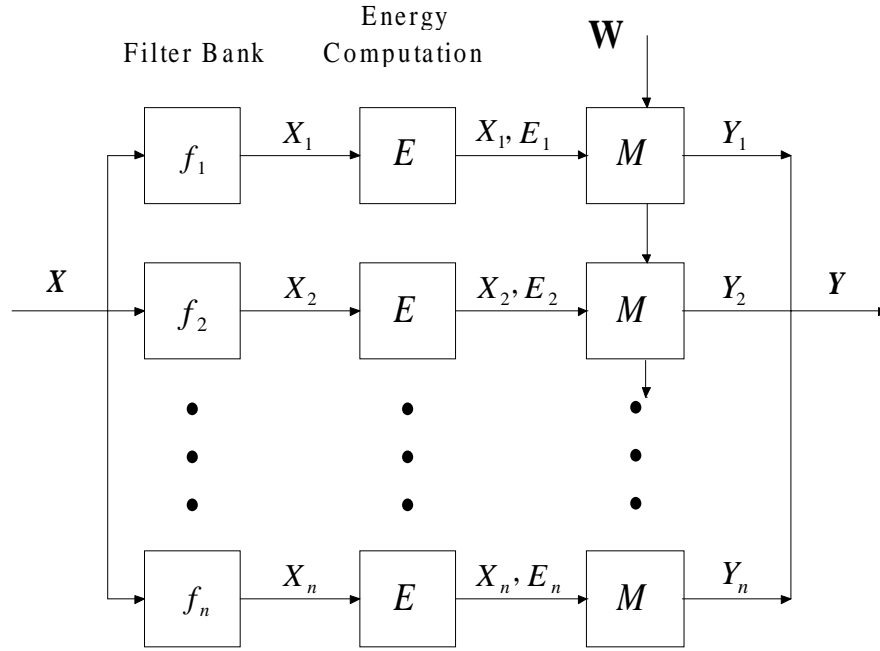

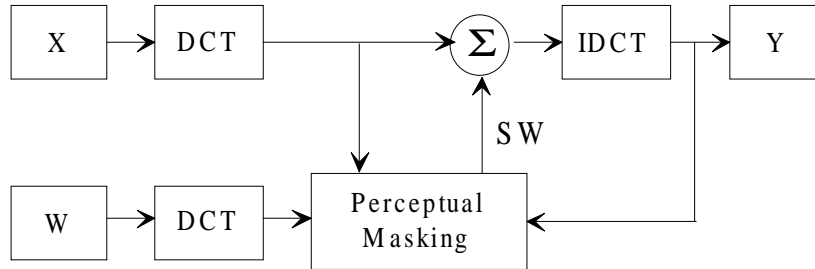Figure 10. Block diagram of sub-band watermarking algorithm.

Figure 11. Block diagram of spectral method with scaling.

The watermark verification procedure is similar to the IA-W method. A hypothesis test is performed on the normalized crosscorrelation coefficient computed between the extracted watermark, W^*, and W. If the result is above a certain threshold, T, the image is authentic. As in previous techniques, T is determined according to the desired probability of detection and

probability of false alarm. An original image and an image watermarked with this algorithm are in Figure 12 and Figure 13; the watermark generated with this technique is shown in Figure 14.[1]
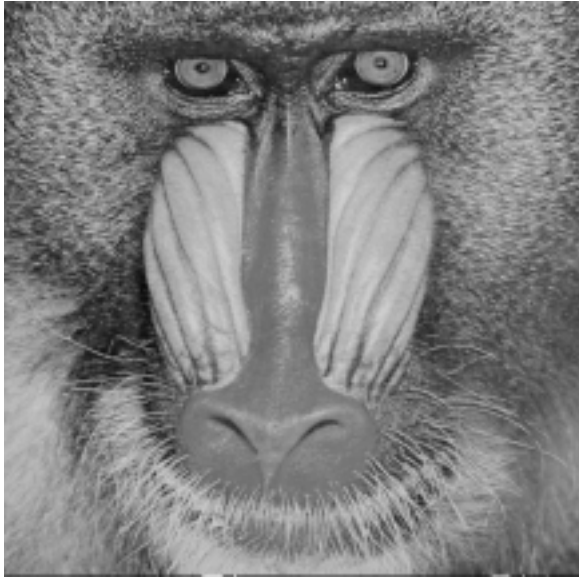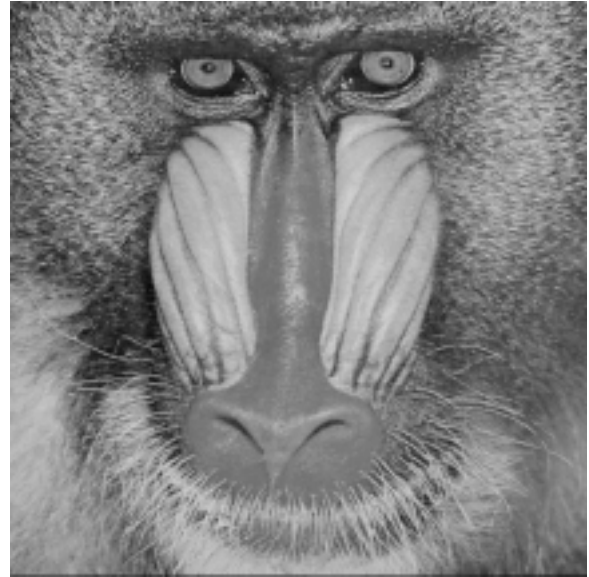

Figure 12. Original Baboon image

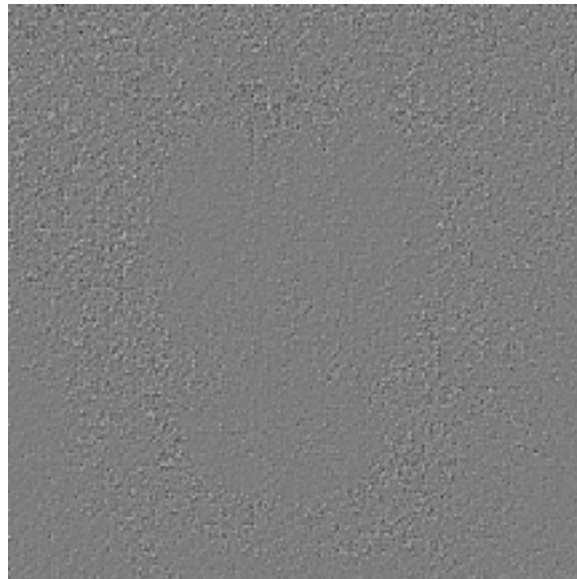Figure 13. Watermarked Baboon image

Figure 14. Watermark for Baboon image

---

[1] These images have been provided by M. Swanson, B. Zhu and A. Tewfik of the University of Minnesota. Used with permission of the authors.

3.5.4 Wavelet-based block-wise scaling of the watermark

A wavelet-based algorithm that also scales the watermark on a block-by-block basis is presented in [74]. *W* for this technique is much smaller than the original image; copies of *W* are tiled throughout the subbands of a wavelet decomposition of *X*. This protects against cropping, but may be susceptible to collusion attacks. *W* is a *2H x 2W* pseudo-random array of ones and negative ones, where *H* and *W* are positive integers. Random numbers from a *N(0,1)* distribution could also be used. The original image *X* (or excerpt of *X* to be marked) should be of size *H2^M* x *W2^M*, where *M* is also a positive integer. This allows for a wavelet decomposition of *X* up to *M* levels. An N_l-level decomposition of *X* is computed. This forms $3N_l + 1$ wavelet subband images. The lowest resolution subband, of size *H2^{M-L}* x *W2^{M-L}*, is not marked. The remaining $3N_l$ subbands are segmented into *H x W* blocks. The *f* [th] frequency orientation at the *l* [th] resolution level is denoted as $X_{l,f}$; the ith block in $X_{l,f}$ is denoted $X_{l,f}(i)$, $l = 1,2,\ldots N_l$ and $f = 1,2,3$. A one-level wavelet decomposition is then computed of *W*. This produces one *H x W* baseband image w_0, and three sub-band images, denoted w_1, w_2 and w_3, each corresponding to a different frequency orientation.

Each block $X_{l,f}(i)$ is marked with a scaled version of w_f, for $f = 1,2,3$.

$$Y_{l,f}(i) = X_{l,f}(i) + \gamma_{l,f}\sqrt{S_{l,f}(i)} \cdot w_f \tag{34}$$

The scale factor, $S_{l,f}(i)$, is defined as the *saliency* of the block, and is a measure of how large the amplitude of the embedded mark can be made while still remaining invisible. $S_{l,f}(i)$ is based on the definition in [75].

$$S_{l,f}(i) = \sum_{u,v} C(u,v)\left|F_{l,f}^i(u,v)\right|^2 \tag{35}$$

C is the contrast sensitivity matrix described in [75], and *u* and *v* are spatial frequency variables (radians per pixel). $F_{l,f}^i$ is the Fourier transform of the block $X_{l,f}(i)$.

$$\gamma_{l,f} = \frac{\alpha E[X]}{\max_i \sqrt{S_{l,f}(i)}} \tag{36}$$

α is a user defined constant adjusted to ensure invisibility of the watermark; $\alpha = 0.1$ was used in [74].

To test an image *Z* with this technique, the same wavelet decomposition used for *X* is computed for Z. The different versions of w_f are extracted from all subbands $Z_{l,f}$. The extracted versions of w_f for each frequency orientation are averaged to form w_f^*. W^* is then constructed from the original w_0, and the extracted w_1^*, w_2^* and w_3^*. A threshold test is then performed on the normalized correlation coefficient between *W* and W^* as in previous techniques.

4   VIDEO WATERMARKS

There has been extensive research done in the area of spatial masking models and applying some of these results to the compression of still images. Developing more sophisticated models, which include temporal masking for the encoding of video sequences, remains an open research area. Some of this is due to the limitations of current video formats. For instance, in order to take advantage of temporal masking effects, we would need to sample the video at a much higher rate than the current 30 frames per second. For this reason much of the work on using visual models for video compression consists of simple rules, such as taking advantage of masking effects at scene changes.

Video watermarking poses some unique requirements beyond those for still image compression because of the additional attacks that video is subject to: frame shuffling, inter-frame collusion, etc. For instance, it is important not to watermark each frame of a video as an independent image. If a different watermark is used for each frame, an attacker can compare frames that change very little within a scene. Simple collusion between these frames can remove a large part of the watermark. Also, the computational effort needed to test each frame in real time would be large. Using the same watermark for each frame also poses problems, since an attacker could then collude with frames from completely different scenes. One method that achieves a tradeoff between marking every frame independently and using one watermark for the entire video is described below. This method - an extension to IA-DCT - marks each I-frame in an MPEG sequence, and allows the motion vectors to carry the mark over to subsequent frames.

4.1   Extension of the IA-DCT Technique to Video

A straightforward extension of the IA-DCT technique discussed here for still images has been extended to video [76]. The just noticeable differences that are based on spatial masking properties do not, however, apply to temporal masking; this means that watermarking each individual frame based on the spatially-derived JNDs will result in visible temporal distortion. A possible way to help reduce this effect within the MPEG framework is to take advantage of the motion vectors, which are available in the MPEG compressed bitstream to propagate the watermark from frame to frame. Although the authors found that this produced visually better results than individual still image watermarking, block artifacts remained in the video sequence. The authors found that the best visual quality was obtained by using the IA-DCT watermarking technique at every I frame and applying a simple linear interpolation of the watermarks to every frame between two consecutive I frames. The overhead necessary to encode the displaced frame difference (DFD) which now consists of the difference between watermarks as well as the displaced signal difference from the original video sequence is negligible, typically adding only 1 to 3 percent in additional data rate. In the watermark encoding process the interpolation technique adds an additional delay consisting of the difference between I frames (typically 15 frames). Several frames of a video sequence that have been encoded using the IA-DCT interpolation method are shown in Figure 15: the top row represents the watermarked frames and the bottom row displays the watermark for each individual frame. Ideally, we would like to use a more formal visual model that takes into account temporal as well as spatial masking in the watermarking process.

4.2   Direct Watermarking of the MPEG-2 Bitstream

Some initial work on applying watermarking schemes to video include the work proposed in [77], where a scheme to embed a digital signature into an MPEG-2 bitstream is presented. The authors propose a watermarking scheme that is similar in spirit to the spread spectrum approach of [46]. For each 8x8 DCT block of the video, a corresponding block is chosen from the watermark signal and transformed using the DCT. The two transformed blocks are added to form the watermarked DCT block. These coefficients are then quantized and Huffman encoded.

The goal is to produce a watermarked video sequence that does not exceed the data rate of the original unwatermarked sequence. The watermarked coefficients are transmitted only if the rate for that coefficient does not increase. The authors claim that typically, 15-30 % of the DCT coefficients is altered. Due to the motion compensated structure of the coder, the authors also address the problem of drift due to the incorrect compensation based on the watermarked blocks. The resulting scheme is of low complexity and allows a watermark signal of several bytes/second.

Figure 15. Video frames watermarked with IA-DCT interpolation

4.3   Scene-Adaptive Video Watermarking

Another proposed method watermarks a video sequence on a scene-by-scene basis [78]. A video is first segmented into scenes, and a *temporal* wavelet transform of each scene is computed. This places the elements of the video that change little over the course of the scene (background, etc.) into the low time-resolution bands of the transform. The parts with fast motion would be decomposed into higher-resolution bands. This effectively isolates the two types of motion in the scene. A different watermark is then placed in each resolution band. This results in the same sequence being embedded in the slow-changing parts of a scene to avoid collusion. The sequences added into the sections with large amounts of motion are different, and change frequently throughout the scene. Different watermarks are placed in each scene (except perhaps for scenes with near-identical backgrounds). The inverse transform of the watermarked wavelet

coefficients forms the marked video sequence. An original video frame, marked frame and the corresponding watermark are shown in Figure 16 through Figure 18.[2]

4.4   Data Hiding

Using the perceptual models, a small data stream may be embedded in an image or video [2,79]. The embedded data is a type of watermark, and the larger image or video that contains the data is called the host object. Perceptual models are necessary to maximize the number of bits a host object can contain, without the embedded data becoming visible. A text file, audio clip, or a low-resolution image or video stream may be inserted into the host depending on the host's size. This technique is referred to as *data hiding*, and can be used to transmit data or authenticate the host object. Authentication is performed through visual (or aural) inspection of the embedded data. Figure 19 and Figure 20 show an original image, and the image with an 8192 bit text file inserted into the original; Figure 21 shows the difference image.[3] If sufficient error correction is used, this data hiding technique can withstand JPEG compression [2]. Another technique that imperceptibly embeds a low-resolution grayscale image into a larger one is described in [80].

5   WATERMARKING STANDARDS

5.1   Standards

Both the MPEG-2 and MPEG-4 video compression standards allow some form of copyright protection to be inserted into the compressed video stream. MPEG-2 does this through the `copyright_extension()` data structure [81]. This data structure, if implemented in the decoder, can be used to prevent decoding and/or copying of the video. Development of a more robust intellectual property identification scheme is underway for MPEG-4 [82,83]. To identify MPEG-4 content, a format for an Intellectual Property Information (IPI) Data Set has been proposed [84]. Each MPEG-4 "object" (video clip, still image, audio clip, text document) would have its own IPI Data Set. This data set can contain the type of data, the owner of the work, time of creation, and a unique identification number for the object. This identifier is to remain in the data stream even if the object is modified. The formal syntax of the IPI Data Set is still under development. The study of intellectual property rights is also part of the developing MPEG-7 standard [85].

There have been several efforts in standardizing copy protection technology. Most of these efforts do not involve the use of watermarking technology but either use encryption techniques or other mechanisms including analog approaches for making images or video either not viewable or recordable. In July 1997 the Data Hiding Subgroup (DHSG) of the Copy Protection Technical Working Group (CPTWG) issued a call for proposals in the area of data hiding and watermarking [86]. Another group that is very interested in the copy protection of images and video is the Copyright Issues Group of the Digital Audio-Visual Council (DAVIC) [87].

---

[2] Different frames from the same sequences have originally appeared in [78]. Used with permission of the authors.
[3] Figure 19 and Figure 20 originally appeared in [2]. All three images are used with permission of the authors.

Figure 16. Original football frame

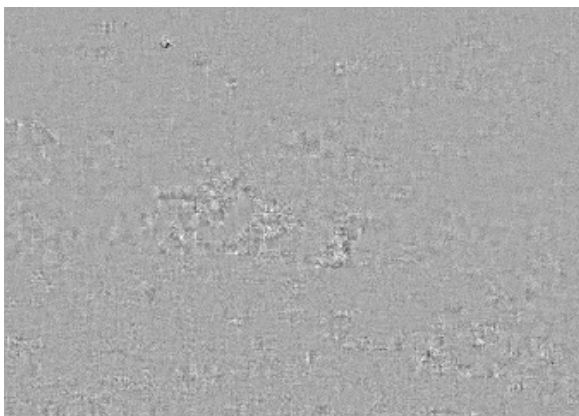Figure 17. Watermarked football frame

Figure 18. Watermark for frame

Figure 19. Original jet image
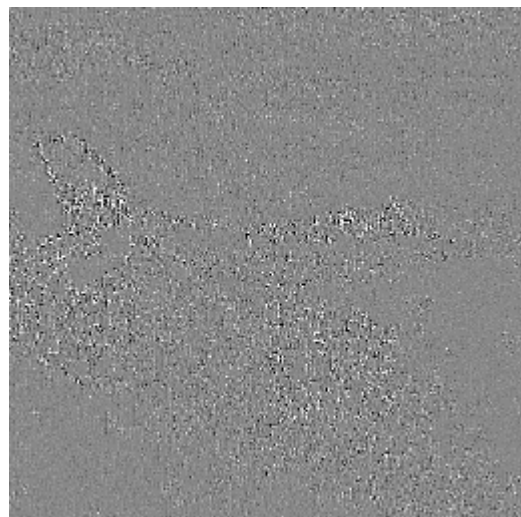

Figure 20. Jet image with embedded text

Figure 21. Difference image

The Digital Versatile Disk (DVD) [88] is the latest technology that has been developed to deliver data to the consumer. One problem that has held up the development of the DVD standard is the protection of copyrighted content [89]. Several media companies initially refused to provide DVD material until the copy protection problem has been addressed. One application that motivated the development of the DVD is the delivery of MPEG-2 encoded full-length movies on a single disk [90]. A 4.7 gigabyte DVD can provide approximately 133 minutes of 4.7 Mb/s MPEG-2 content (3.5 Mb/s for the video, 1.15 Mb/s for the audio and 0.05 Mb/s for miscellaneous information). One way to secure the content on a DVD is to link watermark verification to the proper functioning of the DVD player [91]. For instance, the player's output port would be enabled only upon verification of the watermark. Other methods for copyright protection will result from the DHSG efforts.

6   CONCLUSION

We have described recent developments in the digital watermarking of images and video in which the watermarking technique is invisible and designed to exploit some aspects of the human visual system. Many of these techniques rely either on transparency (low-amplitude) or frequency sensitivity to ensure the mark's invisibility. Watermarks that are image adaptive use visual models to *locally* maximize the amplitude of the embedded watermark, while maintaining the watermark's imperceptibility. This type of algorithm has been shown to be more robust to many attacks based on linear and nonlinear signal processing operations.

The protection of intellectual property rights is perhaps one of the last major barriers to the "digital world." While the techniques presented in this paper are not foolproof, they can help validate claims of ownership that are required for intellectual property law enforcement.

ACKNOWLEDGEMENTS

REFERENCES

[1]   R. B. Wolfgang and E. J. Delp, "Overview of image security techniques with applications in multimedia systems," *Proceedings of the SPIE International Conference on Multimedia Networks: Security, Displays, Terminals, and Gateways*, November 4-5, 1997, Dallas, Texas, vol. 3228, pp. 297-308.

[2]   M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data embedding and watermarking technologies," to appear in *Proceedings of the IEEE*, 1998.

[3]   I. J. Cox and M. L. Miller, "A review of watermarking and the importance of perceptual modeling," *Proceedings of the SPIE International Conference on Human Vision and Electronic Imaging II*, Feb. 10-13, 1997, San Jose, CA, USA, pp. 92-99.

[4] M. Kutter and F. Hartung, "Image watermarking techniques," to appear in *Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information*, 1999.

[5] I. J. Cox, "Spread-spectrum techniques for image watermarking," to appear in *Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information*, 1999.

[6] B. Schneier, *Applied Cryptography*, Second Ed., Wiley & Sons, 1996.

[7] F. Mintzer, G. W. Braudaway and M. Yeung, "Effective and ineffective digital watermarks," *Proceedings of the IEEE International Conference on Image Processing*, October 26-29, 1997, Santa Barbara, CA, vol. 3, pp. 9-12.

[8] W. J. Mitchell, *The Reconfigured Eye: Visual Truth in the Post-photographic Era*, M.I.T. Press, Cambridge, MA, 1992.

[9] D. Stinson, *Cryptography, Theory and Practice*, CRC Press, 1995.

[10] S. Craver, N. Memon, B.-L. Yeo and M. Yeung. "Can invisible watermarks resolve rightful ownerships?" *Proceedings of the SPIE International Conference on Storage and Retrieval for Image and Video Databases V*, Feb. 13-14, 1997, San Jose, CA, vol. 3022, pp. 310-321.

[11] Intellectual Protocols 2, Inc., http://www.ip2.com .

[12] D. Storch and E. Koch, "Controlable user access on multimedia data in the world wide web," *Proceedings of the International Conference on Imaging Science, Systems, and Technology*, June 30 - July 3, 1997, Las Vegas, Nevada, pp. 270-278.

[13] A. B. Watson, "The cortex transform: rapid computation of simulated neural images," *Computer Vision, Graphics, and Image Processing*, 1987, pp. 311-327.

[14] G. E. Legge and J.M. Foley, "Contrast masking in human vision," *Journal of the Optical Society of America*, vol. 70, no. 12, December, 1980, pp. 1458-1471.

[15] A. B. Watson, "Efficiency of an image code based on human vision," *Journal of the Optical Society of America A*, 1987, pp. 2401—2417.

[16] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still Image Data Compression Standard*, Van Nostrand Reinhold Publ., 1993.

[17] H. A. Peterson, A. J. Ahumada Jr. and A. B. Watson, "Improved detection model for DCT coefficient quantization," *Proceedings of the SPIE International Conference on Human Vision, Visual Processing, and Digital Display IV*, vol. 1913, February, 1993, pp. 191-201.

[18] A. B. Watson, "DCT quantization matrices visually optimized for individual images," *Proceedings of the SPIE International Conference on Human Vision, Visual Processing and Digital Display IV*, vol. 1913, February, 1993, pp. 202-216.

[19] R. J. Safranek, "Perceptually based prequantization for image compression," *Proceedings of the SPIE International Conference on Human Vision, Visual Processing, and Digital Display V*, 1994.

[20] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image-dependent quantization and post-quantization data compression," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1989.

[21] I. Hontsch and L. Karam, "APIC: Adaptive perceptual image coding based on sub-band decomposition with locally adaptive perceptual weighting," *Proceedings of the IEEE International Conference on Image Processing*, Santa Barbara, CA, October 26-29, 1997, vol. 1, pp. 37-40.

[22] N. S. Jayant, J. D. Johnston and R. J. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, October 1993, pp. 1385-1422.

[23] B. Zhu and A. H. Tewfik, "Low bit rate near-transparent image coding," *Proceedings of the SPIE International Conference on Wavelet Applications for Dual Use*, vol. 2491, Orlando, FL, 1995, pp. 173-184.

[24] B. Girod, "The information theoretical significance of spatial and temporal masking in video signals," *Proceedings of the SPIE International Conference on Human Vision, Visual Processing, and Digital Display,* vol. 1077, 1989, pp. 178-187.

[25] J.-F. Delaigle, C. De Vleeschouwer, B. Macq, "A psychovisual approach for digital picture watermarking," to appear in the *Journal of Electronic Imaging*.

[26] F. Goffin, J.-F. Delaigle, C. De Vleeschouwer, B. Macq, J.-J. Quisquater, "A low cost perceptive digital picture watermarking method," *Proceedings of the SPIE International Conference on Storage and Retrieval for Image and Video Databases V*, San Jose, CA, Feb. 13-14, 1997, vol. 3022, pp. 264-277.

[27] S. Walton, "Information authentication for a slippery new age," *Dr. Dobbs Journal*, vol. 20, no. 4, April, 1995, pp. 18-26.

[28] R. G. van Schyndel, A. Z. Tirkel, N. Mee, C. F. Osborne, "A digital watermark," *Proceedings of the IEEE International Conference on Image Processing*, November, 1994, Austin, Texas, vol. 2, pp. 86-90.

[29] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison-Wesley, 1995.

[30] U.-C. G. Fiebig and M. Schnell, "Correlation properties of extended m-sequences," *Electronics Letters*, vol. 29, no. 20, September 30, 1993, pp. 1753-1755.

[31] R. B. Wolfgang and E. J. Delp, "A watermark for digital images," *Proceedings of the IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sept. 16-19, 1996, vol. 3, pp. 219-222.

[32] R. B. Wolfgang and E. J. Delp, "Techniques for watermarking digital imagery: further studies", *Proceedings of the International Conference on Imaging Science, Systems, and Technology*, Las Vegas, Nevada, June 30 – July 3, 1997, vol. 1, pp. 279-287.

[33] G. B. Rhoads, *Stenography Methods Employing Embedded Calibration Data*, United States Patent # 5,636,292, June 3, 1997, columns 1-39.

[34] S. Servetto, C. Podilchuk and K. Ramchandran, "Capacity issues in digital watermarking," submitted to the *IEEE 1998 International Conference on Image Processing*.

[35] I. Pitas, "A method for signature casting on digital images," *Proceedings of the IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sept. 16-19, 1996, vol. 3, pp. 215-218.

[36] W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," *Technical Report, Massachusetts Institute of Technology Media Lab*, 1994.

[37] D. J. Fleet and D. J. Heeger, "Embedding invisible information in color images," *Proceedings of the IEEE International Conference on Image Processing*, Santa Barbara, CA, Oct. 26-29, 1997, vol. 1, pp. 532-535.

[38] M. Kutter, F. D. Jordan, F. Bossen, "Digital signature of color images using amplitude modulation," *Proceedings of the SPIE International Conference on Storage and Retrieval for Image and Video Databases V*, San Jose, CA, Feb. 13-14, 1997, vol. 3022, pp. 518-526.

[39] G. W. Braudaway, "Protecting publicly-available images with an invisible image watermark," *Proceedings of the IEEE International Conference on Image Processing*, Santa Barbara, CA, Oct. 26-29, 1997, vol. 1, pp. 524-527.

[40] M. Yeung and F. Mintzer, "An invisible watermarking technique for image verification," *Proceedings of the IEEE International Conference on Image Processing*, Santa Barbara, CA, Oct. 26-29, 1997, vol. 2, pp. 680-683.

[41] H. Kinoshita, "An image digital signature system with ZKIP for the graph isomorphism," *Proceedings of the IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sept. 16-19, 1996, vol. 3, pp. 247-250.

[42] P. Kuosmanen, J. Astola, K. Davíðsson, and K. Halonen, "Digital watermarking through filtering," *Proceedings of the IEEE Nonlinear Signal and Image Processing Workshop*, 1997.

[43] B. Natarajan, "Robust public key watermarking of digital images," *HP Laboratories Technical Report*, no. 97-118, October 1997, pp. 1-10.

[44] K. T. Knox and S. Wang, "Digital watermarks using stochastic screens – a halftoning watermark," *Proceedings of the SPIE International Conference on Storage and Retrieval for Image and Video Databases V*, San Jose, CA, Feb. 13-14, 1997, vol. 3022, pp. 310-316.

[45] G. C. Langelaar, J. van der Lubbe, and R. L. Lagendijk, "Robust labeling methods for copy protection of images", *Proceedings of the SPIE International Conference on Storage and Retrieval for Image and Video Databases V*, San Jose, CA, Feb. 13-14, 1997, vol. 3022, pp. 298-309.

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

[47] F. M. Boland, J. J. K. Ó Ruanaidh and C. Dautzenberg, "Watermarking digital images for copyright protection," *Proceedings of the International Conference on Image Processing and its Applications*, Edinburgh, Scotland, July 1995, pp. 321-326.

[48] A. Piva, M. Barni, F. Bartolini, V. Cappellini, "DCT-based watermark recovering without resorting to the uncorrupted original image," *Proceedings of the IEEE International Conference on Image Processing*, Oct. 26-29, 1997, Santa Barbara, CA, vol. 1, pp. 520-523.

[49] E. Koch and J. Zhao, "Towards robust and hidden image copyright labeling," *Proceedings of the IEEE Workshop on Nonlinear Signal and Image Processing*, Neos Marmaras, Halkidiki, Greece, June 20-22, 1995, pp. 452-455.

[50] J. Zhao and E. Koch, "Embedding robust labels into images for copyright protection," *Proceedings of the KnowRight '95 Conference: Intellectual Property Rights and New Technologies*, 1995, pp. 242-251.

[51] A. G. Bors and I. Pitas, "Image watermarking using DCT domain constraints", *Proceedings of the IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sept. 16-19, 1996, vol. 3, pp. 231-234.

[52] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.

[53] M. Holliman, N. Memon, B.-L. Yeo and M. Yeung, "Adaptive public watermarking of DCT-based compressed images," *Proceedings of the SPIE International Conference on Storage and Retrieval for Image and Video Databases VI*, San Jose, CA, January 28-30, 1998, vol. 3312, pp. 284-295.

[54] Dave Benham, Nasir Memon, Boon-Lock Yeo and Minerva Yeung, "Fast watermarking of DCT-based compressed images," *Proceedings of the International Conference on Imaging Science, Systems, and Technology*, Las Vegas, Nevada, June 30 – July 3, 1997, vol. 1, pp. 243-252.

[55] B. Tao and B. Dickinson, "Adaptive Watermarking in the DCT Domain," *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 21-24, Munich, Germany, vol. 4, pp. 2985-2988.

[56] C.-T. Hsu and J.-L. Wu, "Hidden digital watermarks in images," to appear in the *IEEE Transactions on Image Processing*.

[57] C.-T. Hsu and J.-L. Wu, "Hidden signatures in images," *Proceedings of the IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sept. 16-19, 1996, vol. 3, pp. 223-226.

[58] J. J. K. Ó Ruanaidh, W.J. Dowling and F.M. Boland, "Phase watermarking of digital images," *Proceedings of the IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sept. 16-19, 1996, vol. 3, pp. 239-242.

[59] J. J. K. Ó Ruanaidh and T. Pun, "Rotation, scale and translation invariant digital image watermarking," *Proceedings of the IEEE International Conference on Image Processing*, Santa Barbara, CA, Oct. 26-29, 1997, vol. 1, pp. 536-539.

[60] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989.

[61] K. K. Wong, C.H. Tse, K.S. Ng, T.H. Lee, and L.M. Cheng, "Adaptive watermarking," *IEEE Transactions on Consumer Electronics*, vol. 43, no.4, November 1997, pp. 1003-1009.

[62] X.-G. Xia, C. G. Boncelet, G. R. Arce, "A multiresolution watermark for digital images," *Proceedings of the IEEE International Conference on Image Processing*, Santa Barbara, CA, Oct. 26-29, 1997, vol. 3, pp. 548-551.

[63] C. I. Podilchuk and W. Zeng, "Digital image watermarking using visual models", *Proceedings of the SPIE Conference on Human Vision and Electronic Imaging II*, San Jose, CA, Feb. 10-13, 1997, vol. 3016, pp. 100-111.

[64] C. I. Podilchuk and W. Zeng, "Watermarking of the JPEG bitstream," *Proceedings of the International Conference on Imaging Science, Systems, and Technology*, Las Vegas, Nevada, June 30 – July 3, 1997, pp. 253-260.

[65] C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," to appear in the *IEEE Journal on Selected Areas in Communications*.

[66] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using the wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, 1992, pp. 205-220.

[67] A. B. Watson, G. Y. Yang, J. A. Solomon and J. Villasenor, "Visual thresholds for wavelet quantization error," *Proceedings of the SPIE International Conference on Human Vision and Electronic Imaging*, 1996.

[68] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley & Sons, 1968.

[69] W. Zeng and B. Liu, "On resolving rightful ownerships of digital images by invisible watermarks," *Proceedings of the IEEE International Conference on Image Processing*, Oct. 26-29, 1997, Santa Barbara, CA, vol. 1, pp. 552-555.

[70] A. Papoulis, *Probability, Random Variables and Stochastic Processes,* Third Ed., McGraw-Hill, 1991.

[71] J.-F. Delaigle, C. De Vleeschouwer, and B. Macq, "Digital watermarking," *Proceedings of the SPIE Conference on Optical Security and Counterfeit Deterrence Techniques*, San Jose, CA, Feb. 1996, vol. 2659, pp. 99-110.

[72] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, 1993.

[73] M. D. Swanson, B. Zhu and A. H. Tewfik, "Transparent robust image watermarking," *Proceedings of the 1996 International Conference on Image Processing*, Lausanne, Switzerland, Sept. 16-19, 1996, vol. 3, pp. 211-214.

[74] D. Kundur and D. Hatzinakos, "A robust digital image watermarking method using wavelet-based fusion," *Proceedings of the IEEE International Conference on Image Processing*, Oct. 26-29, 1997, Santa Barbara, CA, vol. 1, pp. 544-547.

[75] T. A. Wilson, S. K. Rogers and L.R. Myers, "Perceptual based hyperspectral image fusion using multiresolution analysis," *Optical Engineering*, vol. 34, no. 11, November, 1995, pp. 3154-3164.

[76] C. Podilchuk, *Lucent Technologies Internal Technical Memo*, Lucent Technologies Bell Labs, 1998.

[77] F. Hartung and B. Girod, "Digital watermarking of MPEG-2 coded video in the bitstream domain", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1997, Munich, Germany, vol. 4, pp. 2621-2624.

[78] M. D. Swanson, B. Zhu, B. Chau and A. H. Tewfik, "Multiresolution video watermarking using perceptual models and scene segmentation," *Proceedings of the IEEE International Conference on Image Processing*, Oct. 26-29, 1997, Santa Barbara, CA, vol. 2, pp. 558-561.

[79] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Data hiding for video-in-video," *Proceedings of the IEEE International Conference on Image Processing*, Oct. 26-29, 1997, Santa Barbara, CA, vol. 2, pp. 676-679.

[80] J. J. Chae and B. S. Manjunath, "A robust embedded data from wavelet coefficients," *Proceedings of the SPIE International Conference on Storage and Retrieval for Image and Video Databases VI*, San Jose, CA, January 28-30, 1998, vol. 3312, pp. 308-317.

[81] B. Haskell, A. Puri and A. Netravali, *Digital Video: an Introduction to MPEG-2*, Chapman and Hall (International Thompson Publishing) New York, 1997, pp. 243.

[82] "MPEG-4 Overview - (San Jose Version): Overview of the MPEG-4 version 1 standard," *International Standards Organization*, Document ISO/IEC JTC1/SC29/WG11 N2078, San Jose, February 1998.

[83] "Overview of MPEG-4 functionalities supported in MPEG-4 version 2," *International Standards Organization*, Document ISO/IEC JTC1/SC29/WG11 N2079, San Jose, February 1998.

[84] "Managing intellectual property identification and protection within MPEG-4," *International Standards Organization*, Document ISO/IEC JTC1/SC29/WG11 N1918, October 1997.

[85] "MPEG-7: context and objectives (version 6 - San Jose)," *International Standards Organization*, Document ISO/IEC JTC1/SC29/WG11 N2082, San Jose, February 1998.

[86] *Call for Proposals*, Data Hiding Subgroup of the Copy Protection Technical Working Group, Version 1.0 July 1, 1997.

[87] *Call for Proposals*, Digital Audio Visual Council (DAVIC). http://www.davic.org .

[88] J. Taylor, *Dvd Demystified : The Guidebook for Dvd-Video and Dvd-Rom*, McGraw Hill Text, 1997.

[89] S. Rupley, "What's holding up DVD?" *PC Magazine*, vol. 15, no. 20, November 19, 1996, pp. 34.

[90] "Leading consumer electronics and entertainment companies propose digital video disc format," *Time Warner Press Release*, Los Angeles, CA, January 24, 1995, http://www.timewarner.com .

[91] I. J. Cox and J.-P. Linnartz, "Public watermarks and resistance to tampering", *Proceedings of the IEEE International Conference on Image Processing*, Oct. 26-29, 1997, Santa Barbara, CA [CDrom version].