Wavelet packets-based Digital Watermarking for Image Verification and Authentication

Alexandre H. Paquet [1],

Department of Electrical and Computer Engineering,
University of British Columbia, 2356 Main Mall,
Vancouver, BC, V6T 1Z4, Canada

Rabab K. Ward,

Institute for Computing, Information and Cognitive Systems,
University of British Columbia, 2366 Main Mall,
Vancouver, BC, V6T 1Z4, Canada

Ioannis Pitas

Department of Informatics,
Aristotle University of Thessaloniki, Box 451,
Thessaliniki 542 06, Greece

Abstract

In this paper, we present a novel watermarking scheme to ensure the authenticity of digital images. Our authentication technique is able to detect malicious tampering of images even if they have been incidentally distorted by basic image processing operations. The image protection is achieved by the insertion of a secret author's identification key in the image wavelet coefficients by their selective quantization. Our system uses characteristics of the human visual system to maximize the embedding weights while keeping good perceptual transparency. We develop an image-dependent method to evaluate, in the wavelet domain, the optimal quantization step allowing the tamper proofing of the image. The nature of multiresolution discrete wavelet decomposition allows the spatial and frequency localization of image tampering. Experimental results are presented to demonstrate the capacity of our system to detect unauthorized modification of images, to show its robustness to image compression and high security levels and to compare it with publicly available image authentication software.

Key words: image authentication, digital watermarking, wavelet packets, tamper proofing, digital security

Contents

List of Tables

Email addresses: alexp@ece.ubc.ca (Alexandre H. Paquet), pitas@zeus.csd.auth.gr (Ioannis Pitas).

URLs: www.ece.ubc.ca/∼alexp (Alexandre H. Paquet), poseidon.csd.auth.gr/EN (Ioannis Pitas).

[1] This research was partially supported by the Natural Sciences and Engineering Research Council of Canada.

List of Figures

1 Introduction

The rapid expansion of the Internet and the overall development of digital technologies in the past years have sharply increased the availability of digital multimedia content. One of the great advantages of digital data is that it can be reproduced without loss of quality. However, it can also be modified easily and, sometimes, unperceptively. In many contexts, such for legal evidence and for video security systems, any modifications of image, video or audio data have to be detected. Therefore, some work needs to be done in order to develop security systems to protect the information contained in digital data.

Watermarking, which allows for the secret embedding of information in a host data, has emerged as a widely accepted approach for copyright protection and ownership identification. A lot of effort has been dedicated to the development of robust watermarking schemes to achieve these goals. Another possible, but less investigated, application of digital watermarking is for the authentication of multimedia content. Fragile watermarking schemes have been developed where any tampering with the image destroys the embedded key. Such systems allows for the detection of unauthorized alterations in an image, a video or an audio sequence. It makes the certification of the validity of digital data possible. As digital media are now widely employed and commonly accepted as official documents, the importance of the protection of their informative content will grow.

In this paper, we introduce a novel technique for content authentication of digital images which is able to detect and localize malicious image alterations while offering a certain degree of robustness to image compression. Our technique introduces a semi-fragile watermarking in an image by the selective rounding of wavelet packets coefficients. This method is superior to previously proposed ones since it allows for the detection of frequency or space domain malicious alterations of an image (compressed or not) without requiring user interaction in the detection process, in which necessary operations can yield major security flaws. In the next section, we review some of the earlier techniques. Then, in section 3, we list the major drawbacks of the previous techniques and explainhow we overcome these problems as we detail the different steps of our semi-fragile watermarking system and highlight the strategy proposed to enhance its security. Finally, we present experimental results in section 4. Conclusions are drawn in section 5.

2 Previous Work

The importance of digital content authentication has made the development of fragile - and semi-fragile - watermarking scheme a hot research topic. Several different approaches have already been proposed. These techniques can be divided into two general categories in terms of the embedding process: acting directly in the spatial domain and others, working in a transform domain. The reader is invited to review more extensive surveys published [2,6,9,20,23]. In particular, [20] gives a generic overview of the watermarking problem while [23] lists the specific requirements of different watermarking systems.

Fragile watermarking techniques that embed hidden information in the spatial domain, such as [3,21,25], are definitely more straightforward and, therefore, less computationally expensive than the ones using some transform domain. Because of the speed advantage, this kind of embedding is probably more suitable for real-time implementation. This is why such techniques have often been extended to the authentication of video data [3]. [25] proposes one of the first watermarking methods for the verification and authentication of high-quality images. A watermark image is invisibly embedded into the source image in the spatial domain by modifying the pixel values. A verification key, which is stored and known only to authorized parties, is produced in the embedding step and used in the verification process to extract the image inserted in the host. However, the embedding process is fragile to unintentional image distortions introduce by basic image processing operations (*e.g.* compression) done for storage purposes.

Another spatial embedding watermarking method is proposed in [21]. Besides allowing the identification of modified regions in tampered images, it is able to reject small distortions introduced by high quality image compression. The watermark is embedded on randomly selected pixels using local neighborhood constraints. The identification of changes or tampering in small details of the image is based on mathematical morphology.

The above methods suffer from the major drawback of spatial domain watermarking: frequency localization of modifications is difficult, if not impossible. Furthermore, most of image authentication systems based on fragile embedding of watermarks in the spatial domain also have the drawback of being more susceptible to malicious attacks. In fact, search -in which the aggressor, who has access to the watermark decoder, creates altered versions of the work and processes them through the decoder by brute force, until one is declared authentic- and collage (collusion) -for which the attacker has access to several similarly marked works and uses parts of different genuine images to form a new authentic image- attacks are a threat to spatial-based, and particularly block-based, watermarking authentication approaches [7]. While search

attacks are more computationally expensive, collage attacks are easily implemented and allow the unregistered modifications of (*supposedly*) tamper-proof images.

The techniques using transform domain are more complex and computationally expensive. Yet, they can offer a higher degree of robustness against common image processing operations [5]. Robustness is important for fragile watermarking systems because it is highly preferable that basic image processing operations - ones that are typically used for the storage of watermarked images - do not alter the embedded marks.

To achieve these goals, some authors have proposed taking advantage of the knowledge of current image compression standards to develop semi fragile watermarking techniques in the DCT domain [12,24]. Although DCT approaches have shown some potential, it is the wavelet domain that has attracted the most attention amongst all the transform domains used. Furthermore, as DCT techniques are mainly block-based, they are also more susceptible to collage attacks. For our application, the main advantage of wavelets over Fourier and DCT analysis is that they allow for both spatial and frequency resolution. Wavelet transform, WT, decomposes a signal in narrow levels of details while keeping the basis signal space limited [8]. This is of great importance when dealing with real signals, especially when spatial localization is to be considered. This explains why WT attracts so much attention for a wide range of image processing applications, including digital watermarking for image authentication [11,13,14,18,26] and the upcoming image compression standard *JPEG-2000*.

A fragile watermarking technique for the tamper proofing of still images is presented in [11]. A mark is embedded in the discrete wavelet domain by the quantization of the image's corresponding wavelet coefficients. The Haar wavelet is used for the image decomposition and a secret identification key is generated by a pseudo-random binary sequence. The rounding of the DWT coefficients to even or odd quantization steps embeds the zeros or ones of the watermark. The embedding locations are stored in the coefficient selection key, *ckey*. In addition, an image-dependent quantization key, *qkey*, is introduced to improve security against forgery and monitor specific changes to the image.

In the same line a digital images authentication procedure that allows for the detection of malicious tampering, while staying robust to incidental distortion introduced by compression is presented in [26]. As in [11], a binary watermark is embedded in the wavelet transform domain. The insertion is again done by the even or odd quantization of selected wavelet coefficients. To augment the robustness of the scheme to image processing operations, the authors propose to make the embedded watermark more robust by rounding the mean value of weighted magnitudes of wavelet coefficients to quantization levels specified

by the predetermined function $Q(x, q)$. The same function is also used in the blind detection process to retrieve the mark privately by reversed quantization. In order to distinguish malicious tampering from incidental distortion, the amount of modification on wavelet coefficients introduced by incidental versus malicious tampering is modeled as Gaussian distributions with small versus large variance. The probability of watermark detection error due to incidental alterations is shown to be smaller than the probability of watermark detection error due to malicious tampering because they produce comparatively smaller variance difference with the embedded marks. The authors argue that this grants a certain degree of robustness to the system and show that their method is able to authenticate JPEG compressed images without any access to the original unmarked image. However, the degree of image compression allowed by the detection procedure is not stated and the selection procedure of quantization parameters is not explained either.

3 Our WP-Based Image Authentication

The two wavelet-based techniques presented above protect digital images from malicious tampering and unauthorized processing while allowing image compression of small ratios. These techniques however require the user to determine the malevolence of an attack. In addition, they necessitate a certain degree of interaction in both the embedding and decoding procedures. Our technique overcomes both these drawbacks as the *accepted attacks* are predetermined prior to the embedding process. Additionally, the embedded parameters are automatically obtained so as to take maximum advantage of the characteristics of the human visual system (HVS). Besides, by using wavelet packets, we further improve on the frequency resolution of the standard discrete wavelet transform. In WP decomposition, not only the output of the low pass filtered image is used for the subsequent decomposition level, but also the high pass filter outputs. This leads to narrower frequency bands at higher frequencies and assures the capture of the image salient points [19]. This also allows for much higher precision and flexibility in the selection of the bands to be used in the embedding process.

Unlike the above mentioned methods, our semi-fragile authentication system does not require the post-processing of the tamper detection in the recovery process. This has the advantage of allowing the original owner or creator of the work - not the end user - to decide the extent to which an alteration is judged acceptable. We believe that techniques including user interaction in the detection process have a serious security flaw. This is why we introduce a novel technique for content authentication of digital images. In this section, we go over the details of the embedding and decoding procedures, highlighting principally the quantization and evaluation steps of each process respectively.

3.1 Embedding Process

The main steps of the embedding procedure developed are presented here and summarized in Figure 2.

1. An author's identification key of 64 bits is produced. The number 64 has been selected since it represents the current standard in identification schemes and is enough for key uniqueness. The author's key is kept secret and is used in the watermark extraction procedure.

2. The key is used to select the decomposition (level and wavelet function) applied to the image. A great advantage of WT is the great flexibility offered by the multitude of basis functions available. In the present application, it increases the security of our scheme, since it is impossible for a would-be pirate to know which specific wavelet domain has been used for the embedding. In our decomposition, we have used either 4 of 5 levels. These numbers were chosen to allow for good frequency resolution and to yield enough bands for embedding while keeping a reasonable computational cost for analysis. The selected wavelets were either Daubechies 12 or 16, Coiflets 12, 28, 24 or 30 [3]. We used those mother functions because while they have finite (compact) support, hence achieve better spatial resolution than most wavelets, they are continuous and yield better frequency resolution than the Haar wavelet. Much work remains in the domain of wavelet function design or selection for particular applications; this is a field of research on its own and the investigation of the effects of wavelet function selection on the final image authentication could be the subject of future work.

3. WP decomposition of the image is performed based on the specification extracted from the key in **2.**

4. The specific wavelet packet coefficients where the mark will be embedded are obtained. Our system identifies 64 groups of $K + 1$ bands. These groups are formed by one principal band surrounded by K secondary bands. The principal band is always located in the top-left corner of the group, that is, it corresponds to the lowest frequencies of the group. The first principal band is selected to be the LL or approximation band, which corresponds to the lowest frequencies of the entire decomposition. Then, the other 63 principal bands are evenly distributed in the wavelet packet decomposition to cover the entire frequency content. In each principal band, our system picks N *subbands* (regions of M wavelet packet coefficients). These subbands are spread along each principal band to cover the spatial content globally at all scales. The position of these regions is fixed in each band for any given decomposition. However, the subbands are shifted (by one subband size) from one main band to the next to cover the entire image content. Each group of bands (a combination of one

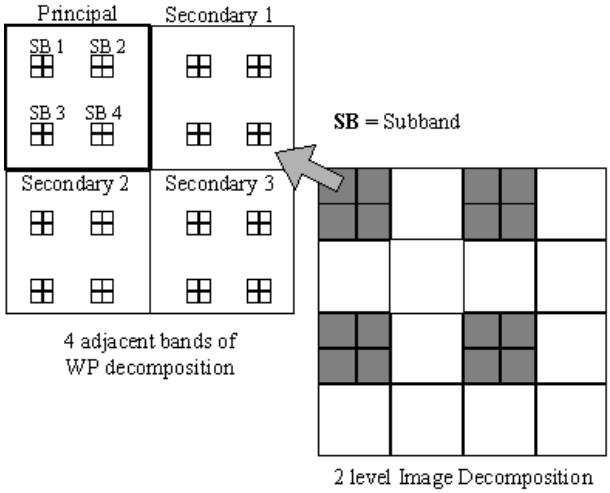


Fig. 1. Coefficients selection approach (step **4**). Four groups of one principal band and three secondary bands are highlighted in gray. These are selected for embedding while the white bands stay untouched. Only four regions (N=4) per band are shown here for clarity.

principal band and K secondary bands) is used in Step **5.** for the embedding of one authentication bit from the author's key. For this reason, N regions of wavelet packet coefficients -or subbands-are established in the secondary bands as well. The regions belonging to a secondary band correspond exactly to the same spatial areas as the WPC regions of the matching principal band. A coefficients selection example is shown in Figure 1 for four groups of four bands, each having four subbands of four coefficients (that is, $K = 3$, $N = 4$ and $M = 4$).

The selection of 64 principal bands with their secondary bands allows for the embedding of the 64-bit author's identification key. This identification step is important because we have to make sure the entire frequency spectrum is covered, as well as the entire image (in the spatial domain), in order to be able to assure authentication of the entire content. Furthermore, since we need to make sure that the images' characteristics are globally protected at each scale, the selected regions have to be spread in space, and their number should be sufficient to cover the image completely. In this context, the application of a four-level WP decomposition and the use of three secondary bands ($K = 3$) for each of the sixty-four principal bands with eight regions ($N = 8$) of four coefficients ($M = 4$) per band, is adequate for the protection of (256 by 256) images [17].

5. The author's secret key is used one last time for its embedding. As mentioned, each of the 64 bits is uniquely embedded in a group of one principal and K secondary bands. First, the means $Mean(i, j)$ of all the selected regions of WPC (in the principal and secondary bands) are individually computed. Then, the original quantization levels $q(i, j)$ are obtained (Equation 1) based

on an optimal quantization step Δ (see Subsection 3.2). Afterwards, each bit of the author's identification key is inserted in the WP domain by the modification of the (individual) mean of the WPC regions belonging to each selected group of bands (one principal and K secondary). Rounding the mean to an even quantization level embeds a zero, while rounding the mean to an odd quantization level embeds a one. This is done by rounding the obtained quantization levels $q(i,j)$ to the nearest even/odd quantization levels (to form $q'(i,j)$) and then adjusting the mean of the WPC regions to the computed values (as shown in Equations 2 and 3). The advantage of using this quantization technique is that the embedded information is modified as a function of the host. By opposition, an adversary can forge a fragile watermark more easily if the embedded pattern is not dependant of the cover work [7]. Therefore, quantization-based approaches increase the security of authentication systems by assuring the uniqueness of the resulting embedded mark.

$$q(i,j) = \lfloor \frac{Mean(i,j)}{\Delta} \rfloor \tag{1}$$

$$\begin{aligned} key[n] = 0 \; q'(i,j) &= \begin{cases} q(i,j) & \text{if } q(i,j) \text{ even} \\ q(i,j)+1 & \text{if } q(i,j) \text{ odd} \end{cases} \\ key[n] = 1 \; q'(i,j) &= \begin{cases} q(i,j)+1 & \text{if } q(i,j) \text{ even} \\ q(i,j) & \text{if } q(i,j) \text{ odd} \end{cases} \end{aligned} \tag{2}$$

$$Mean'(i,j) = q'(i,j) \cdot \Delta \tag{3}$$

In our embedding process, an optimal step Δ is used for the rounding of the mean of the defined regions belonging to the 64 principal bands. On the other hand, $\Delta/2$ is used for the regions belonging to the $64 \cdot K$ secondary bands. This is done to minimize the introduced distortion. Section 3.2 presents the procedure to obtain the optimal Δ for Laplacian distribution of unitary variance ($\sigma^2 = 1$). As the WPC of the selected bands do not have unitary variance, we have to weight each quantization step obtained as a function of the distribution of the particular embedding band in order for it to represent the optimal quantization step. To achieve this, we compute the power (σ^2) contained in each of the selected bands, and use it as the modulating factor. In this way, our system takes advantage of the human visual system's characteristics by giving more weight to marks embedded in regions with more details for which the HVS is less sensitive [10]. In fact, our system takes only implicit advantage of the characteristics of the visual system as it does not-contrary to what others have done [4]-weight the embedded marks as a function of the sensitivity of the human eye at each frequency. However, this has proven to be more than sufficient to assure the invisibility of the marks.

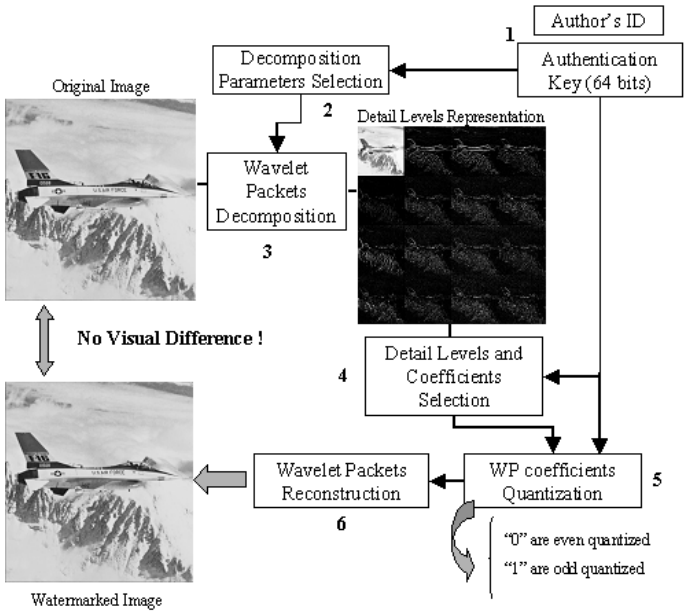


Fig. 2. Embedding Scheme Developed

6. Finally, we apply the appropriate wavelet-packets synthesis bank on the coefficients - some modified and some not - to reconstruct the image which forms the watermarked image.

3.2 Optimal Quantization Step

As we have just explained, our watermarking algorithm uses quantization of selected WP coefficients (Figure 3). The choice of the optimal quantization step Δ is of primary importance in order to maximize the embedding weight (*i.e.* to maximize the watermark importance in the image) while minimizing the distortion introduced in the image by the mark insertion. In terms of coding, it means that we want to minimize the Mean Squared Quantization Error, *MSQE*, which is given by:

$$MSQE = \sigma_q^2 = 2 \sum_{i=1}^{M/2-1} \int_{(i-1)\Delta}^{i\Delta} (x - (\frac{2i-1}{2})\Delta)^2 f_x(x)dx + 2 \int_{(M/2-1)\Delta}^{\infty} (x - (\frac{M-1}{2})\Delta)^2 f_x(x)dx \quad (4)$$

Where f_x is the probability distribution function of the variable x, M the number of quantization levels to be used and Δ is the quantization step selected. The minimization to achieve is therefore:

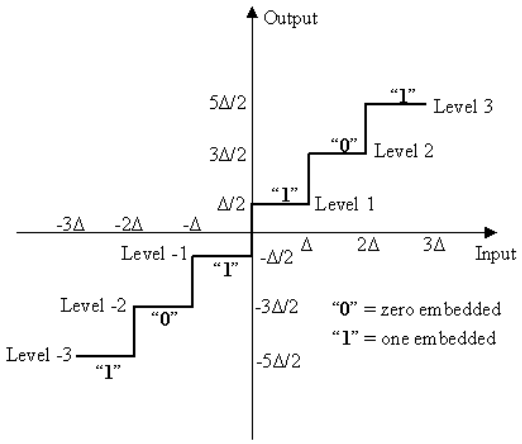


Fig. 3. Input/Output relation in the quantization process

$$\frac{\delta\sigma_q^2}{\delta\Delta} = -\sum_{i=1}^{M/2-1}(2i-1)\int_{(i-1)\Delta}^{i\Delta}(x-(\frac{2i-1}{2})\Delta)f_x(x)dx$$
$$-(M-1)\int_{(M/2-1)\Delta}^{\infty}(x-(\frac{M-1}{2})\Delta)f_x(x)dx = 0 \quad (5)$$

It has been shown that wavelet coefficients have Laplacian distributions [15]. Fortunately, the problem of minimization has already been solved for this type of distribution [1]. Therefore, it means that the optimal quantization steps are already available and that the one to be used can be selected to reflect the degree of protection to be achieved.

3.3 Watermark Decoding Process

The first 4 steps of the decoding procedure are identical to the embedding ones. The author's unique key is first used on the watermarked image to extract a binary verification key without any use of the original *unmarked* image.

5. The coefficients of each subband belonging to each principal band are first scanned. The coefficients of the subbands in the secondary bands are also examined.

6. Intraband comparison: associations are made between the means of the subbands belonging to the same principal band where we embedded the information about one of the bits of the key to decide if the image has suffered from any frequency tampering. Basically, we verify if quantized value of the mean of the coefficients for principal embedding band are of the same parity and if they conform to the corresponding originally embedded bit.

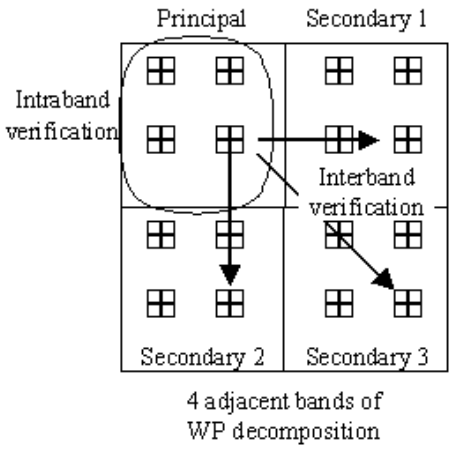


Fig. 4. Intra/interband verification scheme (steps **6.** and **7.**)

7. Interband comparison: verification is now performed throughout the WPC regions associated with the same spatial area to decide whether the image has been spatially altered or not. In this step, we verify if all the bits obtained from the $K+1$ WPC regions (one principal band and K secondary bands) belonging to a given spatial region are of the same parity as the initially inserted bit.

8. Finally, based on the results of steps **5**, **6** and **7**, a decision is made on the authenticity of the image inspected. The rules of detection are quite straightforward and based on experimentations with the embedding of 64 bits in 64 main bands, having 8 subbands of 4 coefficients, and 3 corresponding secondary bands each. First, we state the authenticity of a given spatial region by ensuring that a minimum number of embedded marks (here three regions out of four) corresponding to it are untouched. A similar approach is adopted in the intraband verification; the comparisons are now performed within the same principal decomposition band. The entire FB is declared authentic if at least seven out of eight regions are still marked with the original key. Otherwise, the band is marked as frequency tampered. In addition, we include an overall detection rule that requires more than 504 (out of 512) regions of embedding extracted from the principal bands to have the same quantization level parity as originally embedded for the image to be considered genuine. As each bit is embedded in 8 regions of a unique principal band, the chosen threshold ones assures the detection of one single different bit. Of course, other sets of rules can be defined in order to fully reflect the degree of protection desired. For that reason, they can be considered as the detection parameters to be chosen for each application. Now, we will show how these rules can be used to detect alterations in digital images.

4 Simulation Results

In order to verify the capacities of our watermarking scheme, we have used real and computer generated square-size images to create a set of 720 watermarked images and tested their authenticity. First we have made sure that our embedding system did not introduce visual artefacts in the images to protect and that genuine images were positively authenticated. We also tested our WP-based approach tamper detection abilities. Finally, we compare it with a spatially based publicly available image protection software and examine our system behaviour in the presence of collage attacks.

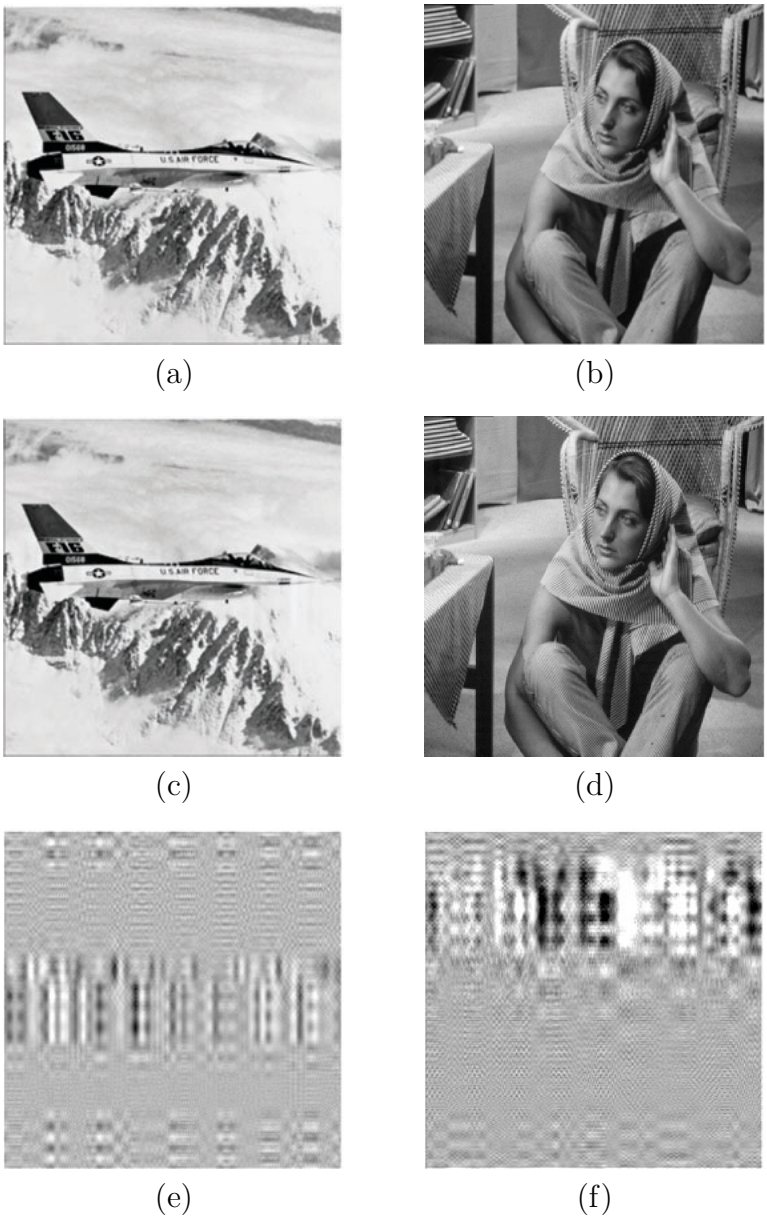
4.1 Embedding, Decoding and Visibility

With the set of images produced, we have obtained an average Peak Signal-to-Noise Ratio (defined by $PSNR(dB) = 10 \cdot \log[\frac{max(I(i,j))^2}{\sum_{N,M}(I^*(i,j)-I(i,j))^2}]$) of 42.47 *dB*. This is above the usually tolerated degradation level of 40 *dB*. Figure 5 shows that the watermarked images are not perceptually different from the original ones and that, in fact, the differences are more evident in the high frequency regions.

Then, we wanted to make sure that untouched images would be declared authentic by our decoding scheme. Using the same set of images, we have the authentication procedure and found that, even if some low-level marks were sometimes destroyed simply by the discretization of the pixel value (*i.e.* saving the image), our intra/interband verification approach was able to declare a genuine image authentic 99.87% of the time. Thus, the false detection rate P_{FR} is 0.13%. In this way, we are assured that the rate of false negatives, which is the number of time where an authentic image is declared tampered, is kept low.

4.2 Tampering Detection

Of course, an important aspect of our system is its ability to localize image tampering. For that reason, we have tampered the previously watermarked *Barbara* image and tested the ability of our system to detect and highlight the doctoring. In addition, since we wanted our system to be robust to high quality JPEG compression, we have compressed the tampered image and ran the verification process again. We found that the ability of our system to detect tampering is excellent (Figure 6), even in the presence of JPEG (or JPEG-2000) compression (Figure 7). However, as the spatial correspondence



(a) (b)

(c) (d)

(e) (f)

Fig. 5. Original (a) and Watermarked (c) (Coifflet 24, $PSNR$= 43.15dB) Airplane Images and (e) the magnified difference between the two. Original (b) and Watermarked (d) (Daubechies 12, $PSNR$= 42.72dB) Barbara Images and (f) the magnified difference between the two.

between WPC and pixels is not one-to-one and also because of the use of regions of WPC, the tampering recognition does not pinpoint specific pixels but defines a region of probable tampering. On the other hand, our system adds the ability to identify tampering, such as filtering, done on the frequency spectrum of a watermarked image.

We have found that, for 256 pixels by 256 pixels grey images, the use of 4 levels decomposition increases the robustness to JPEG compression. This was anticipated since the use of only 4 levels allows for the embedding of

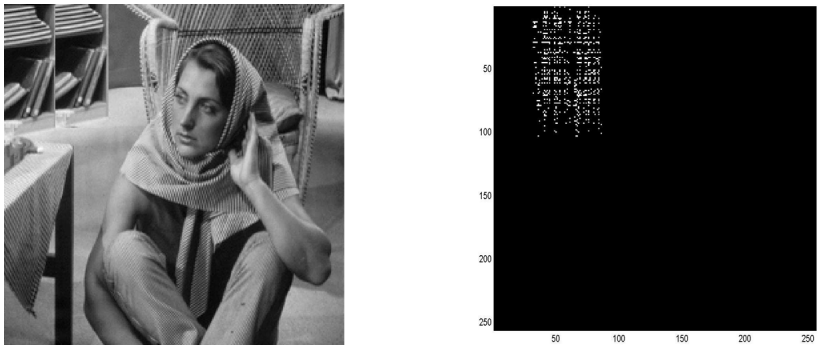


Fig. 6. Tampered Barbara image (a second bookshelf was added to the right of the original one) and spatial tampering detection (Coiflets 24)

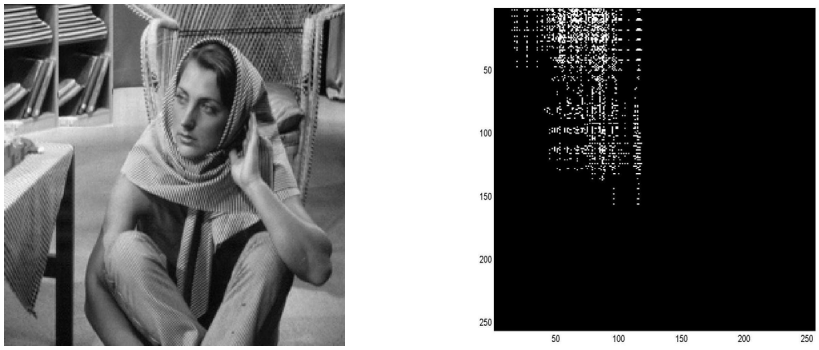


Fig. 7. Compressed (3:1) Tampered Watermarked Image and Detection of spatial tampering (Coiflets 12)

the mark in the most significant parts of the image. As the presence of less decomposition levels generates fewer decomposition bands, it also worsens the frequency resolution. However, this is compensated by better spatial resolution resulting from the fundamental principles of the wavelets. We have also found that the choice of wavelet function for the embedding of the author's mark does not clearly influence the watermarked images quality or the average rate of false negative (see Table 1). However, it affects the tampering detection ability of our system. In fact, we have observed that decompositions based on Coiflets functions yielded better authentication and more accurate spatial localisation of tampering. Intrinsic characteristics of Coiflets wavelets, such as highest number of vanishing moments (both for $\psi(n)$ and $\phi(n)$) and the possibility of perfect reconstruction granted by the non-symmetric nature of the filters [3], allow for better control of the effects of WPC quantization on the resulting images and, therefore, gives more assurance that embedded marks will stay untouched after *discretization* of image intensity. These are, however, highly experimental results. More attention should be given to the design of adapted wavelet functions for tamper proofing of still images. Nevertheless, our wavelet packets-based authentication method has been shown to allow detection and localization of tampering caused by local image modifications, even in the presence of high quality JPEG compression. Subsequently, we shall assess the level of resistance of our watermarking system to collage attacks.

Table 1
Average PSNR and False Negative Rate for different wavelet functions

Wavelet Function	**Average PSNR** *dB*	**Average False Negative Rate** (%)
Coiflets 12	43.40	0.10
Coiflets 18	42.05	0.30
Coiflets 24	41.78	0.15
Coiflets 30	41.98	0.18
Daubechies 12	43.25	0.13
Daubechies 16	43.25	0.03
Overall	42.87	0.13

4.3 Comparison with another authentication scheme

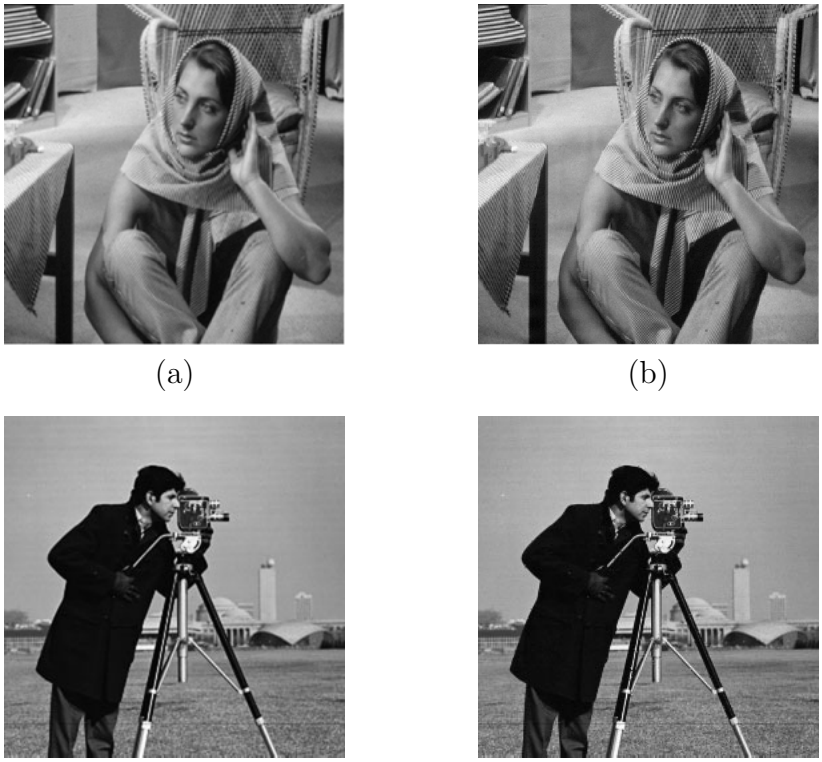
As *copyright protection* is the most common application of digital watermarking, some efforts have been put forth for the development of testing benchmarks for *robust* watermarking schemes. *Stirmark* [2], *Certimark* [3] and *Optimark* [4] are examples of systems that try to solve problems associated with the evaluation and comparison of watermarking systems intended to protect copyright of digital works. Nevertheless, all those benchmarks are meant to compare robust watermarking schemes only.

In the absence of evaluation standards for fragile watermarking systems, we need to compare our scheme with previously proposed ones on a one on one basis. For that reason, we have chosen a commercially available watermarking software called *Eikonamark*. It is a crude implementation of [3], and comes from the *Alpha-tec* corporation [5]. Using only the image authentication capabilities of this software (which also offers robust embedding for copyright protection), we compared the quality of the produced images, the ability to detect tampering, and finally, the resistance to collage attacks achieved by this software with the results obtained with our system.

[2] http://www.cl.cam.ac.uk/ fapp2/watermarking/stirmark/
[3] http://www.igd.fraunhofer.de/igd-a8/projects/certimark/
[4] http://poseidon.csd.auth.gr/optimark/
[5] http://www.alphatecltd.com/watermarking/

4.3.1 Image Quality and Tampering Detection

As the first comparison step, we produced watermarked images with the spatial-based scheme and compared them with the ones obtained with our novel approach. Figure 8 shows original images in comparison with their marked version. From the comparison of the average *PSNR* value obtained with this software (38.42 *dB*), with the one obtained with our system (42.62 *dB*), it is clear that the commercial software degrades the images more than our method does.



(a) (b)

(c) (d)

Fig. 8. Original (a) and Spatially-marked -*PSNR*= 38.51 *dB*- (b) Barbara images and Original (c) and Spatially-marked -*PSNR*= 38.37 *dB*- (d) Cameraman images

Another important characteristic is the ability of the systems to correctly identify modified regions in marked images. To assess this for the spatial-based scheme, we have performed the same doctoring tests as the ones presented in Section 4.2.

The tampering detection quality achieved here (Figure 9) is comparable with the results obtained with our system (shown in Figure 6). However, as this system is spatially based, it yields a slightly better delimitation of the modified areas. On the other hand, we found that it is possible for doctored regions to go unnoticed in several cases. First, if the tampering of a 256 by 256 gray scale image is smaller than 10 by 10 pixels, we have found that Eïkonamark is not able to detect the tampering. At a size of 10 by 10 pixels, the doctoring is

Fig. 9. Tampered Spatially Marked Barbara Image and Detection

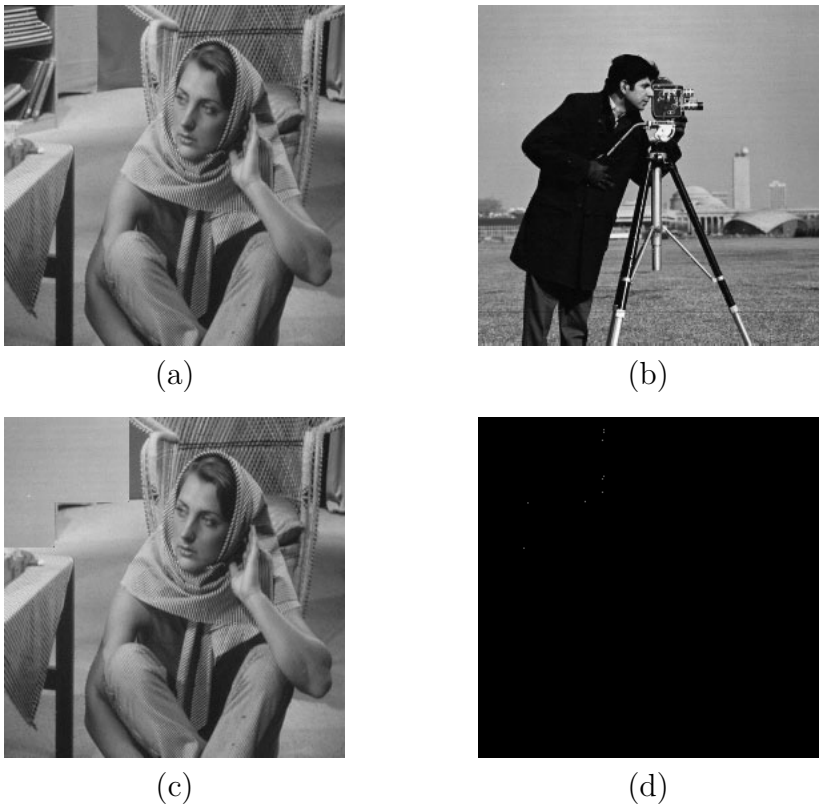
detected in the sense that the authentication key is not 100% found, but the tampering cannot be localized. In fact, if the size of the region of tampering is 10 by X (or X by 10), where $X \leq 30$, the watermark is not noticeably broken, and the modified area cannot be spatially pin pointed. Moreover, the publicly available scheme might generate a somewhat better detection of spatial tampering for *medium-size* corrupted areas, but it does not allow for the localization of frequency tampering. In addition, we have found the system to be highly susceptible to high quality JPEG compression, as shown in Figure 10.

Fig. 10. Compressed (3:1) Spatially Marked Barbara Image and Authenticity Detection

4.3.2 Resistance to Collage Attacks

Since the goal of authentication methods is to detect any unlawful modification or tampering, it is of utmost importance to be able to uncover any kind of alterations. Forged attacks are often problematical since they are designed to bypass implemented protections. In that sense, collage attacks -the combination of several marked images to form another, tampered image- are easy to realize and have been proven effective at defeating authentication procedures. For that reason, the last aspect of our comparison considers the capacity of the marked images to resist collage attacks, or more precisely, the ability of the authentication processes to detect them. For our system and the spatial-based one, we first obtained two watermarked images. Then, we produced a

tampered image by using parts of the two *authentic* images. Finally, tamper detection results are presented for both approaches.



(a) (b)

(c) (d)

Fig. 11. Spatially Marked Barbara (a) and Cameraman (b) Images with the Mixed Version (c) , notice the disappearance of books from top left corner, and the Tampering Detection Result with Commercial Software (d)

From Figure 11, it is obvious that the spatial-based system is easy to defeat by the use of collage attack. Even if the attack is made quite apparent for visualization reasons, the spatial-based approach is not able to detect the combination of two original images. Of course, this is a major security flaw as collage attacks are easily implemented and are quite effective at removing (or adding) important visual information. By comparison, our system is able to declare the image *tampered*. As shown in 12, our WP-based approach is not able to locate the changes due to the high level of modifications detected. In fact, in this case, the localization of the doctoring is not of utmost importance as the image inspected is formed by the combination of two genuine images. Therefore, no region can be considered more or less tampered than the others. It is the image, as a whole, that needs to be considered unauthentic. From this, the results obtained are obviously satisfactory, as we have shown that the proposed method is more secure than the commercially available software. In addition, as stated earlier, straightforward spatial embedding eases search attacks, while the embedding in a wavelet domain-unknown to potential attackers-prevents it or, at least, makes it substantially more lengthy and difficult.

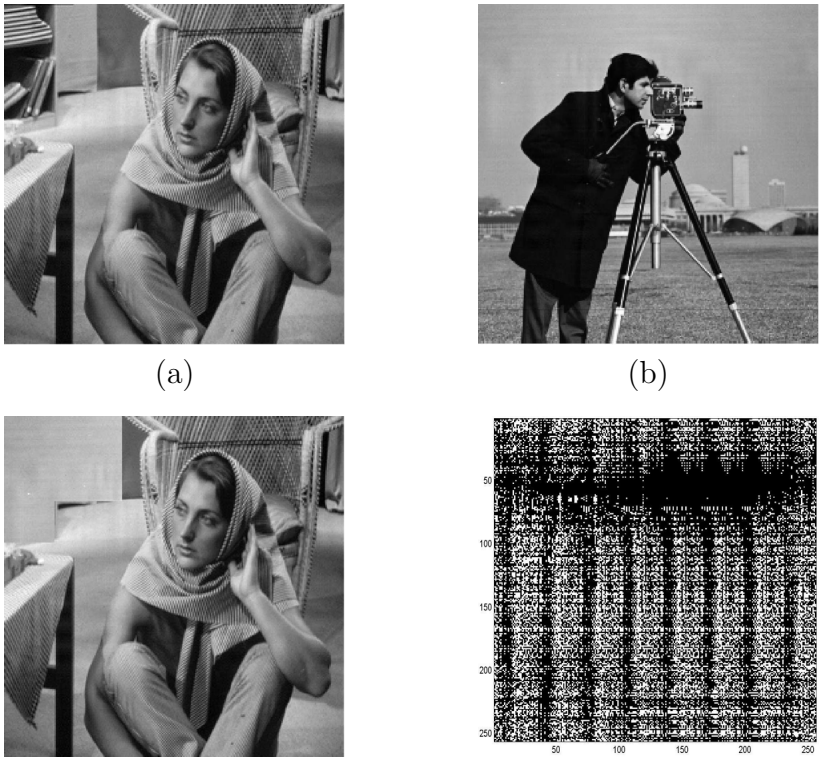


Fig. 12. Watermarked Barbara (a) and Cameraman (b) Images with the altered version (c), and the tampering detection result with our WP-based approach (d)

4.3.3 Summary of Comparisons

To sum up, we have demonstrated that our system introduces less visual distortion when embedding marks in gray-scale images. We have shown that the quality of spatial tampering detection is equivalent for both systems, with a slight advantage, as far as localization of large tampered regions, for Eïkonamark. However, our system includes the recognition of frequency tampering in addition to spatial ones. Moreover, it surpasses the commercial watermarking software in its capacity to tolerate image processing operations since it already includes robustness to high quality JPEG compression. Furthermore, our system is able to detect collage attacks. In conclusion, our system outperforms the spatial-based system in most of the aspects investigated. With minor changes and optimization, it can certainly be developed into commercial software.

5 Conclusions and Future Work

In this paper, we have presented a technique for digital image authentication based on semi-fragile watermarking. Our method embeds an author's secret

identification key in the image to protect its authenticity by quantizing selected wavelet packets coefficients. The system proposed is highly secure, as only the original owner of the watermarked work knows the specific domain of embedding. We developed an optimal quantization step evaluation procedure to take advantage of the HVS characteristics. This allows us to maximize the mark embedding weights while minimizing the distortion introduced. By showing that the watermarked images are visually identical to their *unmarked* originals, experimental results have demonstrated the effectiveness of this approach. Furthermore, it has been shown that our intraband/interband verification technique allows good spatial and frequency localization of tampering without requiring access to the original image. Besides, the rate of false negative detection achieved is low. We have also found that the use of wavelet packets-based embedding domain maximizes the robustness of the marks, which allows our system to work in the presence of high quality JPEG compression. Finally, we have proven that our semi-fragile watermarking technique can detect collage attacks while showing the advantages of our method compared with a commercially available authentication scheme.

The proposed technique shows excellent promise as it allow the authentication of digital images without preventing their efficient storage. Nonetheless, there are still many aspects that could be further investigated. For example, coding techniques could be used to increase the embedding capacity. The addition of an error control-coding module to augment the reliability of the information carried would achieve this goal [16]. Furthermore, the use of chaotic mixing in the watermark generation would add a supplementary layer of security to our system [22]. Future work could also include extending its resistance to JPEG and/or JPEG-2000 compressions by the extraction of compression-invariant image characteristics in the wavelets domain and their use in the embedding process. Finally, as wavelet packets decompositions are not limited to two-dimensional signals, the concepts developed for the authentication of images could be adapted on other media such as audio or video.

References

[1] W.C. Adams and C.E. Giesler. Quantizing characteristics for signals having laplacian amplitude probability density function. *IEEE Transactions on Communications*, 26(8):1295–1297, 1978.

[2] M. Barni, C.I. Podilchuk, F. Bartolini, and E.J. Delp. Watermark embedding: Hiding a signal within a cover image. *Special Issue of IEEE Communication Magazine on Digital Watermarking for Copyright Protection: A Communication Perspective*, 39(8):102–108, August 2001.

[3] F. Bartollini, A. Tefas, M. Barni, and I. Pitas. Image authentication techniques

for surveillance applications. *Proceedings of the IEEE*, 89(10):1403–1418, October 2001.

[4] S. Bhattacharjee and M. Kutter. Compression tolerant image authentication. In *IEEE International Conference on Image Processing (ICIP'1998)*, volume I, pages 435–439, October 1998.

[5] I.J. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for images, audio, and video. In *IEEE International Conference on Image Processing (ICIP'96)*, volume III, pages 243–246, 1996.

[6] I.J. Cox, M.L. Miller, and J.A. Bloom. Watermarking applications and their properties. In *IEEE International Conference on Information Technology*, pages 6–10, 2000.

[7] I.J Cox, M.L. Miller, and J.A. Bloom. *Digital Watermarking*. Morgan Kaufmann Publishers, 2929 Campus Drive, Suite 260, San Mateo, CA 94403, USA, 2002.

[8] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992. Notes from the 1990 CBMS-NSF Conference on Wavelets and Applications at Lowell, MA.

[9] F. Hartung and M. Kutter. Multimedia watermarking techniques. *Proceedings of the IEEE*, 87(7):1079–1107, July 1999.

[10] N. Jayant, J. Johnston, and R. Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10):1385–1422, October 1993.

[11] D. Kundur and D. Hatzinakos. Digital watermarking for telltale tamper proofing and authentication. *Proceedings of the IEEE*, 87(7):1167–1180, July 1999.

[12] C. Lin and S. Chang. Semi-fragile watermarking for authenticating jpeg visual content. In *SPIE Security and Watermarking of Multimedia Content II*, pages 140–151, San Jose, CA, January 2000.

[13] C.-S. Lu and H.-Y.M. Liao. Multipurpose watermarking for image authentication and protection. *IEEE Transactions on Image Processing*, 10(10):1579–1592, October 2001.

[14] A. Lumini and D. Maio. Blind watermarking system for digital images in the wavelet domain. In *SPIE International Symposiom Electronic Imaging Security and Watermarking of Multimedia Contents II*, pages 524–535, January 2000.

[15] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[16] L. M. Marvel and C. T. Retter. The use of side information in image steganography. In *IEEE International Symposium on Information Theory and Its Applications (ISITA'2000)*, November 2000.

[17] A.H. Paquet. Wavelet packets-based digital watermarking for image authentication. Master's thesis, University of British Columbia, Vancouver, B.C., July 2002.

[18] A.H. Paquet and R.K. Ward. Wavelet-based digital watermarking for image authentication. In *IEEE Canadian Conference on Electrical and Computer Engineering*, volume I, pages 879–884, Winnipeg, Manitoba, May 2002.

[19] A.H. Paquet, S. Zahir, and R.K. Ward. Wavelet packets-based image retrieval. In *IEEE International Conference on Acoustics Speech and Signal Processing*, volume IV, pages 3640–3643, May 2002.

[20] C.I. Podilchuk and E.J. Delp. Digital watermarking: Algorithms and applications. *IEEE Signal Processing Magazine*, 18(4):33–46, July 2001.

[21] M.A. Tefas and I. Pitas. Image authentication and tamper proofing using mathematical morphology. In *European Signal Processing Conference (EUSIPCO'2000)*, September 2000.

[22] M.A. Tefas and I. Pitas. Image authentication based on chaotic mixing. In *IEEE International Symposium on Circuits and Systems (ISCAS'2000)*, volume I, pages 216–219, May 2000.

[23] G. Voyatzis and I. Pitas. The use of watermarks in the protection of digital multimedia products. *Proceedings of the IEEE*, 87(7):1197–1207, July 1999.

[24] M. Wu and B. Liu. Watermarking for image authentication. In *IEEE International Conference on Image Processing (ICIP'1998)*, volume II, pages 437–441, October 1998.

[25] M.M. Yeung and F. Mintzer. An invisible watermarking techniques for image verification. In *IEEE International Conference on Image Processing (ICIP'1997)*, volume II, pages 680–683, 1997.

[26] G.J. Yu, C.-S. Lu, H.-Y. M. Liao, and J.-P. Sheu. Mean quantization blind watermarking for image authentication. In *IEEE International Conference on Image Processing (ICIP'2000)*, volume III, pages 706–709, Vancouver, BC, Canada, 2000.