Coefficient Selection Methods for Scalable Spread Spectrum Watermarking

Angela Piper[1]*, Reihaneh Safavi-Naini[1], and Alfred Mertins[2]

[1] School of Information Technology and Computer Science
University of Wollongong, NSW, Australia
[2] Faculty of Mathematics and Science
University of Oldenburg, Oldenburg, Germany

Abstract. With the increasing use of mobile and internet technology as a means of content distribution, we see a move towards scalable content distribution mechanisms, which allow the adaptation of content to the requirements of the target system. Adequate protection of scalable content requires the adoption of scalable watermarking algorithms. Many of the existing algorithms are based on the spread spectrum techniques, first presented in a watermarking context by Cox et al. [2]. The various algorithms often use quite different methods for coefficient selection, resulting in different watermark properties. We examine the effect of the coefficient selection method on watermark scalability by considering the quality and resolution scalability of seven simple selection methods when used as part of a spread spectrum watermarking algorithm.

1 Introduction

The internet is already seeing widespread use as a mechanism for content distribution. A growing number of devices are being connected via networks, all with differing capabilities and requirements in areas such as resolution, colour depth, storage capacity and processing ability. Furthermore, the users of these devices may also have differing needs or desires with regards to the display of such content. Finally, in order to provide delivery within acceptable time it is desirable that the transmitted content be as small as possible. As a result, in order to best distribute content over a network it is becoming increasingly necessary to tailor the content to the requirements of both the device and the user. That is, the content must be highly scalable.

Already there is a great deal of discussion about how to best protect the rights of the creator, owner or distributor of an image. Digital watermarking has the ability to provide such protection as well as offering a variety of other potential uses. However the process of tailoring the image to different device and user requirements means that the majority of devices will receive only parts of the image content, and different devices will receive different parts. This in

---

* Partial funding for this research was provided by the Smart Internet Technology Cooperative Research Centre, Australia.

turn means that the majority of devices will receive only parts of the watermark and that different devices will receive different parts of the watermark. Thus, it is necessary to ensure that our watermarking algorithms are scalable, because each of these different devices must still be able to obtain sufficient data from its partial watermark to allow it to detect the embedded information at the required level.

Many of the existing watermarking algorithms are based on spread spectrum techniques. One of the components of any spread spectrum watermarking algorithm is its coefficient selection method. There are a seemingly endless variety of coefficient selection methods available, each with its own motivation, strengths and weaknesses. We consider seven simple coefficient selection methods and evaluate their scalability, both in terms of quality and resolution scalability.

2 Background

2.1 Spread Spectrum

Spread Spectrum modulates the signal with a pseudo-random noise sequence to produce the watermark X. Insertion requires both an embedding formula and a method for coefficient selection, and there are many options in each area.

Three embedding formulae are provided in [2]:

$$v_i' = v_i + \alpha x_i \tag{1}$$
$$v_i' = v_i(1 + \alpha x_i) \tag{2}$$
$$v_i' = v_i(e^{\alpha x_i}) \tag{3}$$

where α is a (potentially variable) scaling factor used to ensure visual undetectability and x_i is the ith element of X.

Although Cox et al offer only one coefficient selection method, the 1000 largest non-DC coefficients in the greyscale component, numerous other options have been implemented including non-LL subband wavelet coefficients which exceed a given significance threshold [3], blue component coefficients only [4] or coefficients in all three colour channels [7].

The detection of a spread spectrum watermark is achieved by examination of a correlation coefficient. In the case where the original image is available, the embedding process can be reversed and the correlation between the candidate and extracted marks can be calculated. A large correlation value corresponds to watermark presence and a small value indicates that the candidate watermark is absent from the image.

2.2 Scalability in Compression

A scalable image compression algorithm is one which allows an image to be compressed for a number of target bit rates such that an optimal image can be reconstructed, at any of those rates, using the relevant sections of the same compressed data.

There are two main types of scalability to consider in the case of still images: resolution scalability and quality scalability.

Resolution scalability (or *spatial scalability*) is achieved by encoding a low resolution version of the image separately from one or more layers of higher resolution data. This data can be combined with the appropriately scaled low resolution version to produce a higher resolution image. Typically each refinement-layer allows the display of an image at twice the horizontal and twice the vertical resolution previously obtainable.

Quality scalability is achieved by encoding a coarsely quantised version of the image separately from one or more layers of more finely quantised refinement data at the same resolution. The refinement-layers can be combined with the coarsely quantised version of the image to produce a higher quality image. Quality scalability is also termed *SNR scalability*, however the quality metric used to determine the layers need not be directly related to the signal-to-noise ratio (SNR).

The extent to which content is scalable depends on the number of refinement-layers. If few layers of refinement data are used then the resulting compressed bit stream will be optimal for only a few target bit rates. A larger number of refinement-layers will provide optimality for further bit rates, creating a stream with higher scalability at the cost of a slightly greater total length.

2.3 JPEG2000

JPEG2000 is a new wavelet-based image compression standard which has been developed to provide higher levels of consistency, performance and flexibility than the old DCT-based JPEG standard. An important feature of the standard as it applies to internet and mobile applications is that JPEG2000 offers greatly improved options for scalability.

As part of the compression process, a discrete wavelet transform is applied to decompose the image into four subbands: LL, LH, HL and HH. The LH, HL and HH subbands form the highest resolution layer. The LL subband can be further decomposed using the same procedure, and the resultant LH, HL and HH subbands form the second highest resolution layer. The decomposition process is continued until all desired resolution layers are obtained; the final LL subband forms the lowest resolution layer.

A context adaptive bit plane coder is independently applied to groups of coefficients from the same subband to produce a number of coding passes. These passes can then be arranged into quality layers so that those passes which provide the greatest amount of information about the image are in the lowest layer, those which provide slightly less information appear in the next layer and so on. Precisely how many passes are assigned to each layer will depend upon the compression rate set for that layer.

3 Scalability in Watermarking

3.1 Previous Work

The concept of scalable watermarking was first introduced by Wang and Kuo [10] as the watermarking component of an "integrated progressive image coding and watermarking system", allowing simultaneous image protection and image compression with progressive display. They provide no formal definition however, either here or in later work with Su[9].

Chen and Chen [1] explicitly emphasize that upon receipt of "more information of the watermarked image, the bit error rate (BER) of the retrieved watermark image decreases". They add that such an algorithm must "take into consideration the way the image is transmitted", tailoring the algorithm to the scalable transmission method rather than simply progressively transmitting a watermarked image.

The watermarking scheme of Steinder et al. [8] restricts the watermark to the base-layer only in the interests of early detection. While this will certainly provide early detection, all refinement layers remain unprotected by the watermark.

These papers focus on early detection as opposed to detection under a variety of possible rate constraints, which may be the reason that it is not until the discussion by Lin et al. [5] that explicit mention is made of the requirement implicit in [8] that the watermark be "detectable when only the base-layer is decoded".

3.2 Proposed Definition

The purpose of scalable watermarking is to suitably protect content regardless of which particular portions are delivered. Clearly then, such algorithms must protect the content in its base form, but the extra commercial value contained within the higher layers of the content warrants a greater amount of protection. Thus we define a scalable watermarking algorithm as follows:

A scalable watermarking algorithm is a combined watermark embedding and detection scheme intended for use with scalable content and possessing the following two properties:

1. The watermark is detectable in any portion of the scaled content which is of 'acceptable' quality.
2. Increased portions of the scaled content provide reduced error in watermark detection.

In this definition we do not include the point made by Chen and Chen that the watermark should be tailored to the scalable coding or transmission method because, while this may well be necessary to achieve the above properties, should it be possible to achieve the properties without such tailoring, the algorithm would still be well suited to the outlined purpose.

3.3 Applying the Definition

The proposed definition is useful in a conceptual sense; however in order to perform a meaningful evaluation of any watermarking schemes in light of this definition it is necessary to convert the qualitative terms into quantitative ones. As is the case with any such conversion, the particular selections made will be substantially application dependent. A consequence of this is that in any general study, such as this one, the choices made will always be somewhat arbitrary.

If the watermark is to be deemed detectable in any portion of the scaled content which is of acceptable quality, we require definitions for what constitutes *detection* and for what constitutes *acceptable* quality. The point at which a watermark is considered detectable will depend both on what rates of error are considered acceptable for the specific application and on the accuracy of the model of the detection statistic used to estimate the appropriate detection threshold for a given probability of false positive error. For this study we will follow Cox et al. in employing the similarity statistic as our measure of correlation, considering the candidate vector detectable if the similarity between it and the extracted vector exceeds the detection threshold corresponding to a false positive probability of 1×10^{-9}. Rather than employing the constant threshold used in [2] we apply the more accurate false positive model proposed by Miller and Bloom [6] which increases towards that proposed by Cox et al. as the size of the extracted watermark increases.

Precisely what constitutes acceptable quality is highly subjective. The author or distributor of the content will generally make some assessment as to the level of degradation that the content can be expected to sustain before it no longer holds either commercial value or artistic merit. Given that scalable encoding involves the selection of a base-layer or lowest quality version of the image, we can reasonably assume that this base-layer constitutes the least acceptable quality version. Thus the smallest portion of acceptable quality we might want to consider would be an image composed solely of the lowest quality or lowest resolution layer.

We also wish to ensure that increased portions of the scaled content provide *reduced detection error*. Although it is possible to obtain a far finer granularity, we can still obtain an accurate picture of a watermarking scheme's general behaviour by defining an *increased portion* of the content as a full added quality or resolution layer. The *error rate* is the proportion of the total detection attempts which are either false positive, where we detect a watermark which is not in fact present, or false negative, where we fail to detect a watermark which is present. It is not possible to obtain accurate estimates of the error rates for these systems using only a few trials. However, provided the shape of the distribution of the detection statistic does not change significantly, an *increase in average similarity value* will correspond to a *reduction in error rate*. So, rather than attempt to measure the error rate directly, we will consider the mean similarity value, for which it is far easier to obtain an accurate estimate.

Even with this defined, there still remains the problem as to what sort of increase in similarity is desirable. If we wish to ensure that all portions of the content are equally protected then we would require that equal amounts of the watermark vector be detectable from each single portion of the content. If this is the case, then the ideal detector response from an image constructed from the first k of n layers would be $\sqrt{\frac{kN}{n}}$ where N is the length of the watermark vector. However there is no particular reason for treating all resolution or quality layers uniformly. It is quite likely that particular resolution or quality layers contribute far more to the overall image than do others. Thus it might be preferable to employ some measure of the value of a given layer in determining what amount of the watermark should be used for the protection of that layer. In order to do this we would take a perceptual distortion measure D, such as the peak signal to noise ratio (PSNR), and determine the desired similarity based on the reduction in distortion provided by each layer. For example, if the PSNR for an image reconstructed from the first layer lies halfway between that of a mid-grey image[3] and that of an image reconstructed using the first and second layers, then we would want an equal amount of the watermark to be embedded in each of these layers. We would, of course, want the similarity value of an image composed of all n layers to equal the expected similarity value for the full length N watermark. Thus, if $D(k)$ is the distortion between the original image and the image reconstructed using the first k of n layers, the ideal detector response from the reconstructed image would be $\sqrt{\frac{(D(k)-D(0))N}{(D(n)-D(0))}}$, where the 0th layer consists of a mid-grey image.[4]

4 Coefficient Selection Methods

Now that we have outlined the desired behaviour of a scalable spread spectrum watermark we consider different methods for coefficient selection and investigate which, if any, match that desired behaviour. There are numerous possibilities for selecting the coefficients in which to embed and almost every watermark algorithm will use a different scheme. The following selection methods are an attempt to provide a variety of schemes, with some justification for each.

top: Embed in the 1000 largest magnitude coefficients, regardless of any other considerations. This scheme has the greatest freedom of coefficient selection and thus the most intrinsic robustness that can be provided when embedding proportionally to coefficient magnitude. However because there is no restriction on embedding in either the chrominance (C_b and C_r) components or the low resolution subbands, it risks visual detectability unless the embedding strength α is low.

---

[3] A mid-grey image is our best reconstruction based on no information.

[4] It should be noted that this calculation relies on the ability of the distortion measure to accurately reflect the relative visual quality of two images. If the distortion values obtained do not lie in the range where this is the case then the use of this calculation as an ideal value becomes suspect and an alternative should be found.

nolow: Embed in the 1000 largest magnitude coefficients, excluding the lowest resolution layer. The lowest frequency subband is often excluded from watermark embedding schemes due to the sensitivity of the human visual system to artifacts caused by the modification of these bands.

lum: Embed in the luminance component only. Many spread spectrum watermarking schemes are designed for greyscale images only. Embedding in the luminance (Y) component only avoids the risk of detectability through variations in colour, which can be quite apparent at strengths where variations in brightness are not detected. However this restricts the selection space to one third of the available coefficients.

lumnl: Embed in the luminance component only, excluding the lowest resolution layer. This is perhaps the scheme closest to that recommended for colour images in [2] and can be expected to share the advantages and disadvantages of both lum and nolow.

res: Embed in each resolution layer proportionally to the number of coefficients in that resolution. The number of coefficients added by the second resolution is three times that available at the first resolution and each subsequent addition provides four times that provided by the one before it. Furthermore, the sensitivity to modifications in each resolution is reduced as the resolution layer increases. Thus we can comfortably embed an increasing portion of the watermark in each additional resolution whilst maintaining quite a high embedding strength.

comp: Embed in each component proportionally to the number of coefficients in that component. This scheme allows embedding in colour components, which are commercially valuable and may warrant such additional protection, but it ensures that only one third of the watermark is embedded in any component in an attempt to avoid colour artifacts due to excessive embedding in a single component. However in images where colour coefficients are not large, this is likely to embed more of the watermark in the colour components than does the unconstrained embedding.

top2/5: Embed in those coefficients with magnitude greater than two fifths of the largest coefficient in their associated resolution layer. This selects coefficients which are fairly large for their resolution layer but has less emphasis on the lower resolution layers (which tend to have higher valued coefficients) than the unconstrained scheme. Strategies involving variable thresholds such as this generally do not specify a set watermark length, however for comparison purposes we maintain a length of 1000 and stop embedding once that length is reached.

5 Experimental Setup

To investigate the effects of the above mentioned coefficient selection methods on watermark scalability, the following experiment is performed:

We examine the three classic 512 × 512 RGB test images: lena, mandrill and peppers. Each image undergoes JPEG2000 compression using 6 resolution layers, precincts of size 128 × 128, and quality layers with rates 0.01, 0.02, 0.04, 0.06 and 0.9999. A transformation from RGB to YC_bC_r space is applied, as is usual, to decorrelate the colour channels and thus achieve better compression results. Given that the addition of a watermark will cause some degradation to the image there is no reason to employ a lossless filter. Thus the wavelet transformation uses the Daubechies 9,7 filter, as this is the lossy filter provided by the core of the JPEG2000 standard.

Directly preceding the quantisation stage of compression, a spread spectrum watermark of length 1000 is embedded into the wavelet domain coefficients. Following [2] we use Gaussian pseudo random noise with zero mean and unit variance, a single bit message, and

$$v_i' = v_i(1 + \alpha x_i)$$

as the embedding formula. In order to provide consistent grounds for comparison, the scalar embedding strength α is adjusted for each selection method and each image in order to ensure that the mean squared error of the resulting image is 6.5[5] given the full resolution and a rate of 0.9999.

Once an image has been watermarked and compressed we can produce from it a series of images which comprise the first k layers of the full image, where k ranges from 1 to 6 for a decomposition by resolution or from 1 to 5 for a decomposition by quality. These images represent what might be received by various devices with different resolutions or bandwidth. The watermark X' can be extracted from any of these images V' by taking it and the unwatermarked image V and applying the inversion of the embedding formula. The corresponding similarity value is then calculated using

$$sim(X, X') = \frac{X.X'}{\sqrt{X'.X'}}$$

To obtain a more accurate estimate of this value we perform the above procedure using 100 different Gaussian watermarks and record the average similarity value for each k-layer image. In each case, a threshold is calculated based upon the length of the extracted watermark, using the false positive model[6] described in [6], and whether or not the similarity value passes this threshold is recorded.

6 Experimental Results

6.1 Adjusted Embedding Strengths

We see low embedding strengths for top and comp, a somewhat higher embedding strength for lum, and much higher strengths for nolow, lumnl and res. The top and comp schemes are completely unconstrained with regards to resolution

---

[5] A mean squared error of 6.5 corresponds closely to a peak signal to noise ratio of 40, which should ensure visual undetectability.

[6] This model is adapted to work with similarity rather than normalised correlation

and select a large number of coefficients from the lowest resolution layer, thus requiring a low embedding strength to maintain a consistent level of distortion. While lum is also unconstrained with regards to resolution, the restriction to a single component ensures that no more than one third of the coefficients in the lowest resolution layer are available for selecting, thus it is impossible for the lum scheme to select as many lowest resolution coefficients as do top and comp, hence it is likely to produce lower distortion and allow a higher embedding strength. Interestingly, the top2/5 scheme shows embedding strengths very near to those of top and comp, showing that this scheme still selects a high number of the more visually detectable coefficients. The nolow, lumnl and res schemes, which are drastically restricted in their ability to select low resolution coefficients, all achieve a resultant increase in embedding strength.

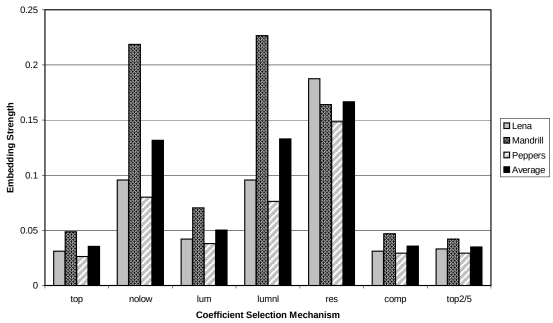

Fig. 1. Embedding Strengths

6.2 Detectability

We now examine whether or not the watermarks established using these selection methods are detectable in the scaled content. Given that we perform only a hundred trials on each image, we would not expect to see any missed detections for a scheme which achieves a reasonably low false negative error rate. Thus we consider the resolution and the quality layer at which each selection scheme first passes the detection threshold, on all trials and for all three images [Tab. 1].

Table 1. Layer at which detection threshold is exceeded for all images

	top	nolow	lum	lumnl	res	comp	top2/5
Resolution	1	5	5	5	4	1	1
Quality	2	1	2	1	1	3	2

Unfortunately, no scheme survives at both the lowest quality layer and the lowest resolution layer of all images. That is, none of the schemes examined here fully satisfy the first property of a scalable watermark: that the watermark be detectable in any portion of the content which is of acceptable quality. We cannot expect the nolow, lumnl and res schemes to be detectable in the lowest resolution layer, due to the resolution restrictions placed on these schemes, however undetectability until layer 4 is unacceptable. The detection results for decomposition by quality layer are reasonably good, although the low embedding strength schemes fail detection at the first layer. The two schemes which provide the best detectability overall are the top2/5 and top schemes, both of which are fully detectable at the lowest resolution layer and at the second quality layer.

6.3 Decreasing Error Rate - Resolution

For each selection method, we can compare the similarity value obtained from the image reconstructed upon receipt of resolution layer k with an ideal similarity value. As was discussed in Section 3.3, we calculate the ideal similarity value based on the reduction in distortion provided by each layer as $\sqrt{\frac{(D(k)-D(0))*1000}{(D(6)-D(0))}}$, where $D(k)$ is the distortion, in this case the PSNR, between the original image and the image reconstructed using the first k resolution layers.

It can be easily seen [Tab. 2] that none of the schemes provides a consistently close match to our ideal for all three images. The nolow, lumnl and res schemes are inherently disadvantaged due to the impossibility of obtaining close to the ideal value at the first resolution layer, and these schemes continue to remain well under the ideal until the final layers have been added. The top and comp schemes tend towards the opposite extreme, embedding too much of the watermark in the lowest layer and not enough in the final layers. This problem is less severe than that encountered by the highly constrained schemes and the match for top and comp is best on average. The moderately constrained schemes, top2/5 and lum, suffer in the average case from their poor performance on the mandrill image.

Table 2. Average squared deviation from the ideal - resolution

	top	nolow	lum	lumnl	res	comp	top2/5
Lena	9.37	67.29	7.06	67.30	184.12	31.70	3.40
Mandrill	23.47	123.37	45.87	129.59	77.88	3.18	36.73
Peppers	4.18	86.95	7.80	72.24	172.59	8.03	6.24
Average	12.34	92.53	20.24	89.71	144.86	14.30	15.46

6.4 Decreasing Error Rate - Quality

The same examination can be performed using a quality decomposition. Again, similarity values obtained from the image reconstructed upon receipt of quality layer k are compared with an ideal similarity value based on the reduction in distortion provided by layer k.

As was the case with the resolution decomposition, there is no scheme which is closest to the ideal for all three images. Furthermore, while for the resolution decomposition the less constrained schemes were generally close to the ideal and the more constrained schemes were generally far from the ideal, there is no such consistency to be found in the quality decomposition.

Table 3. Average squared deviation from the ideal - Quality

	top	nolow	lum	lumnl	res	comp	top2/5
Lena	57.05	17.63	63.59	17.76	27.05	103.54	70.30
Mandrill	33.32	41.96	19.19	43.62	33.63	61.77	15.10
Peppers	106.28	21.56	101.82	29.74	10.65	141.16	96.91
Average	65.55	27.05	61.53	30.37	23.78	102.16	60.77

As was the case with detectability at low layers, the scalability of a given selection method with respect to error reduction favours, on average, the nolow, lumnl and res schemes under a quality decomposition. The schemes which do not have high embedding strengths, perform poorly in terms of quality scalability. However, even the highest strength schemes do not fit the ideal well and are outperformed by both lum and top2/5 on the mandrill image.

The most striking feature of the quality scalability results, however, is the exceedingly large deviation for the comp scheme. It seems that with only the most significant bits of the colour coefficients being assigned to low quality layers, the similarity values for comp are always far below the ideal, much more so than the other schemes which have much the same embedding strength but are free to select a higher number of coefficients from the luminance channel.

7 Conclusion

None of the selection methods examined in this experiment fully provide the properties of a scalable watermarking algorithm. The first property we require is that the watermark be detectable in any acceptable portion of the image. Unfortunately, while the high embedding strengths achievable using selection methods that provide minimal distortion (nolow, lumnl and res) allow watermark detectability in the lowest quality layer, we are not able to consistently detect a watermark embedded using such schemes in images of low or moderate resolution. Conversely, those selection methods which allow consistent detectability when an image is adapted for low resolution display (top, comp, and top2/5) require low embedding strengths and are not consistently detectable in our least acceptable quality image. The lum selection method, which allows embedding at a moderate strength, is consistently detectable at neither the base resolution nor the base quality layer. The top scheme, which selects the largest coefficients across all resolutions, and the top2/5 scheme, which selects the large coefficients within each resolution, have the best general results, both allowing detection at the lowest resolution layer and the second quality layer.

The same problems occur with reducing error rate as increased portions are received, and no selection scheme performs consistently well in this regard. The top, comp, top2/5 and, to a lesser extent, lum schemes provide quite a close match to our ideal rate of error reduction as resolution layers are added. However these schemes, particularly comp, deviate highly from our ideal during receipt of the quality scalable bit stream. The nolow, lumnl and res selection methods are closest to our ideal rate of error reduction as quality layers are added, but are far from the ideal in terms of resolution scalability.

Given the conflict between watermark scalability in terms of resolution and watermark scalability in terms of quality, it seems unlikely that a fully scalable watermarking algorithm can be achieved merely by altering the coefficient selection method. Instead, the better schemes, such as top and top2/5, should be used as a foundation from which to investigate whether added scalability might be achieved through alternate embedding formulae and variable strength embedding.

References

1. T. P.-C. Chen and T. Chen, "Progressive Image Watermarking"; Proc. IEEE Intl. Conf. on Multimedia and Expo., July 2000.
2. I. J. Cox, F. T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia" IEEE Trans. Image Processing, 1997.
3. R. Dugad et al., "A New Wavelet-Based Scheme for Watermarking Images"; International Conference on Image Processing, vol 2, 1998.
4. Kutter et al., "Towards Second Generation Watermarking Schemes"; International Conference on Image Processing, 1999.
5. E. Lin, C. Podilchuk, T. Kalker, and E. Delp, "Streaming video and rate scalable compression: What are the challenges for watermarking?", Proceedings of the SPIE Security and Watermarking of Multimedia Contents III, January 2001
6. M.L. Miller and J.A. Bloom, "Computing the Probability of False Watermark Detection"; Proc. 3rd International Workshop on Information Hiding, 1999.
7. A. Piva et al., "Exploiting the cross-correlation of RGB-channels for robust watermarking of color images"; International Conference on Image Processing, vol 1, 1999.
8. M. Steinder, S. Iren and P. D. Amer, "Progressively Authenticated Transmission"; MILCOM, November 1999.
9. P.-C. Su, H.-J. M. Wang and C.-C. J. Kuo, "An Integrated Approach to Image Watermarking and JPEG2000 compression"; Journal of VLSI Signal Processing, vol. 27, 2001.
10. H.-J. M. Wang and C.-C. J. Kuo, "An Integrated Progressive Image Coding and Watermarking System"; Proc. Int. Conf. on Acoustics, Speech and Signal Processing, vol. 6, pages 3721–3724, March 1998.