Content Based Watermarking of Images

Mohan S Kankanhalli
School of Computing
National University of Singapore, Singapore 119260
Email: mohan@iscs.nus.edu.sg

Rajmohan[*]
K R Ramakrishnan
Department of Electrical Engineering
Indian Institute of Science, Bangalore 560012
Email: krr@ee.iisc.ernet.in

Abstract

With the rapid growth of networked multimedia data systems, copyright protection of proprietary multimedia work has gained importance. Inserting a robust and invisible signal (*watermark*) that clearly identifies the owner or the recipient is beginning to emerge as the solution. We present a novel invisible and robust watermarking technique for images that can be easily extended for video data. Previous watermarking research have only partially used the results of the human visual system (HVS) studies done to evaluate the JPEG quantization table. This also does not provide a framework for the spatial domain watermarking methods. We propose a new way of analyzing the noise sensitivity of every pixel based on the local region image content, such as texture, edge and luminance information. This results in a *just noticeable distortion* mask for the image to be watermarked. Then each bit of the watermark is spread spatially and shaped by a pseudo-noise sequence such that its amplitude is kept below the noise sensitivity of the pixel into which it is embedded. It can be either embedded in the spatial domain or can be DCT coded to be embedded in the transform domain. Experimental results show that the resultant watermark is resistant to various attacks such as JPEG compression, cropping, addition of noise and is perceptually invisible.

1 Introduction

Digitizing of multimedia data has had a dual effect. While on one hand it has enabled faster and more efficient signal storage, transfer and processing, on the other hand duplication and manipulation of such a signal has also become very easy and undetectable. Security concerns over copyright violation of multimedia data have also increased with the growth of computer networks that enable fast and error free movement of any unauthorized duplicate and possibly manipulated copy of multimedia information. The only solution appears to be to cement into the image, video or audio data a secondary signal that is not perceivable and is bonded so well with the original data that it is inseparable and survives any kind of multimedia signal processing. Techniques to embed/retrieve such a stamp or secondary information (*watermark*), that conveys some information about the intended recipient or the lawful owner of the original data, have been of considerable research interest. Work in this area is challenging as it demands a good understanding of multimedia signal processing, communication theory and the Human Visual/Audio System (HVS) to create a watermark that is not perceivable (*transparent*) yet unaffected (*robust*) by signal processing methods that attempt to separate or remove the embedded bits without perceivably degrading the original data quality. There has also been an increase in commercial activity in this area as seen by the popularity of software products such as "PictureMark" [3].

Image watermarking techniques proposed so far can be broadly divided into spatial domain and transform domain methods, They can further be classified based on whether or not the original unwatermarked image is used in the watermark extraction process. Using the original image in the extraction process provides greater strength to the embedded bits but restricts the use of such a watermark. An embedding technique, that ensures retrieval without the original is therefore desired for wider applications such as embedding captions, annotations etc. Watermarking techniques exploit the redundancies in the coded image to embed the information bits. Redundancies in images could be statistical or perceptual. We will discuss the discrete cosine transform (DCT) coded image due to its use in the JPEG standard. By transforming spatial data into another domain (such as spatial frequency), statistical independence between pixels as well as high-energy compaction can be obtained. The general method of DCT coding involves dividing the original spatial image into smaller $N \times N$ blocks of pixels, and then transforming the blocks to obtain equal-sized blocks of transform coefficients in the frequency domain. These coefficients are then thresholded and quantized in order to remove subjective redundancy in the image. Results obtained in the study of the sensitivity of the HVS to the DCT basis images to generate the default JPEG quantization table [16] have been extensively used to decide the location and the strength of the watermark.

---

[*] Presently with the Corps of Signals, Indian Army

1.1 Attacks on Watermarks

A watermarked image is likely to be subjected to certain manipulations, some unintentional such as compression and transmission noise and some intentional such as cropping, filtering etc. These manipulations are also called *attacks* [2, 9]. Some of the attacks that could be thought of are as under:

1. Geometrical distortions. Rotation by an integer or non-integer value, spatial scaling, cropping of an area of the image, translations, cut and paste to another background/image.
2. Addition of a constant offset to the pixel values.
3. Addition of Gaussian or non Gaussian noise.
4. Linear filtering such as low pass or high pass filtering.
5. Non linear filtering such as median filtering.
6. Lossy compression.
7. Local exchange of pixels.
8. Quantization and requantization.
9. D to A and A to D conversion.
10. Watermarking of watermarked images. The original watermark should be still correctly recovered or multiple watermarking should render the image useless.
11. Dithering distortion.
12. Color reduction.
13. Printing the image and rescanning / photocopying.
14. *Collusion.* A number of authorized recipients of the image should not be able to collude and use the differently watermarked copies (such as by averaging out) to generate an unwatermarked copy of the image.
15. *Forgery.* A number of authorized recipients of the image should not be able to collude to form a copy of watermarked image with the valid embedded watermark of a person not in the group with an intention of framing a third party.
16. *Counterfeiting.* When there is a need of the original in the decoding process, it should not be possible to produce a fake original that also performs as well as the original and also results in the extraction of the watermark as claimed by the holder of the fake original. The counterfeiting scheme works by first creating a counterfeit watermarked copy from the genuine watermarked copy by effectively inverting the genuine watermark. This inversion creates a counterfeit of the original image which on comparison of the decoded versions of both the original and counterfeit original yields the forged signature(inverted one). Thus the technique of establishing watermarked image with the original one breaks down.

1.2 Previous Work

A number of spatial domain methods have been proposed in the past but we are not aware of the use of HVS models to arrive at the distortion tolerance of every pixel in the image and its use to vary the change at every pixel location. Kutter et. al [7] ensure that the strength of the watermark at any position is a percentage below the luminance values. Since luminance is not the only factor that contribute to the sensitivity of the eye, we feel that such a watermark is not as robust it could have been. Hartung et. al [4] have proposed a technique similar to spread spectrum communication. The method though very robust and perceptually superior, does not exploit the subjective redundancy in the image/video frame. The watermark is uniformly embedded. To ensure that such a watermark is not visible the amplitude of the watermark has to be kept below the most sensitive threshold. This therefore could be appreciably improved by varying the strength depending on the underlying content of the image. Langellar et. al [1] and Delp et. al [13] also suffer from this drawback.

Many transform domain methods proposed so far have emphasized the importance of varying the watermark strength in different DCT coefficients. Cox et. al [5] make the N highest valued DCT coefficients candidates for watermarking and the strength is kept proportional to the value of the coefficient. Podilchuk et. al [10] use the visual model developed by Watson [17]. The alteration to any basis image is kept below the distortion tolerance level of the DCT basis function so as to keep it invisible. Ruanaidh et. al [6] use the JPEG quantization table to weigh the DCT coefficients and compare the energy content contribution of each coefficient to the total energy. The coefficient is watermarked only if the coefficient has a significant contribution. Tao et. al [14] have an approach that is similar to that of ours. Here the block as a whole is given a sensitivity label that shapes the watermark. But their method is not extendible to the spatial domain and embedding is done only in the transform domain.

We further observe that though the HVS models have been utilized in DCT domain methods of watermarking, using the quantization matrix provided in the matrix in the JPEG standard can only work well at a global level since the quantization matrix cannot be changed within an image. The use of the DC coefficient of the DCT block in such cases, to cater for the luminance sensitivity also does not guarantee a true representation as it is only the average of the block. It may work for blocks with low variance but may deviate substantially in certain image blocks. For example a checker board pattern of black and white will be assessed as mid-gray. We propose a content based watermarking technique that assigns to each pixel position a just noticeable distortion (JND) level based on the local image content. The content parameters used are texture, edge and luminance information. We have applied this to the basic watermarking technique of [4] and show that it improves both the robustness and the perceived quality of the watermarked image. We show that this method is resistant to attacks such as JPEG compression, cropping and addition of noise. However, since we have applied this on the spatial-domain watermarking technique of [4], our watermarking algorithm shares its weakness i.e. it is not very robust against rotation and scaling attacks. But it must be noted that our method is a general technique in the sense that it can be used to improve any spatial watermarking technique. Therefore its role is complementary with respect to all spatial-domain watermarking techniques. For example, the PictureMark software [3] uses a spatial domain technique which alters the values of those image pixels which possess relative extrema values in the local neighborhood [11]. However, the amount of alteration is fixed for the entire image. This technique can be improved by doing the alteration adaptively by using the content-based JND mask generated by our technique. Another advantage of our method is that it can be easily extended to the compressed domain by DCT coding the JND mask and then applying it to the compressed image.

The content based classification of the image will be discussed in section 2. In section 3 the classification will be utilized in a spread spectrum like watermarking technique for adaptive embedding and extraction of watermark bits. The experimental results will be given in section 4. The advantages of the proposed method and our areas of future work will be summarized in the concluding section.

2 Content based Image Segmentation.

Watermarking an image is essentially the process of altering the pixel values of an image in a manner that ensures that a viewer of the image does not notice any perceptual change between the original and the watermarked image. Since any alteration of the image pixel values could be treated as a form of noise, we will interchangeably use the term distortion and alteration to imply the process of embedding or watermarking. Altering a large number of pixel values in an arbitrary manner will result in noticeable artifacts. Every pixel value of an image can be altered only to a certain limit without making perceptible difference to the image quality. This limit can be called as the *just noticeable distortion* or *JND* level. We will call an array of JND values of the size of original image as a JND mask. This mask expresses a critical distortion profile in the sense that if the distortion caused by the watermarking algorithm is at or below the thresholds at all points in space the degradation in the original image quality is imperceptible. We propose a new algorithm to generate the JND mask from the image as shown in fig 1 and fig 2. This mask is then utilized to embed the watermark in the image. Our watermark is therefore guaranteed to be imperceptible and since we embed till the JND limit the watermark is also robust. The robustness is further enhanced by embedding the same bit of information into a number of image pixel values.

The visibility of distortion in a region of the image depends on the underlying image content features as listed below:

- **Edges.** Edge information of an image is the most important factor for our perception of the image. It is in fact the necessary and sufficient information that is to be transmitted if the final receiver is the HVS [15]. It has the least noise sensitivity i.e. the lowest just noticeable distortion value and it is essential to maintain edge integrity so as to preserve the image quality.
- **Smooth Areas.** Psycho-visual studies [8][15] have shown that the HVS has a general bandpass characteristic. Smooth areas influence our perception together with the edge information. The JND perception thresholds are relatively low as compared to strongly textured regions.
- **Textures.** The distortion visibility is low when the background has strong texture [8]. A strongly textured region has a very high noise-sensitivity level.
- **Brightness Sensitivity.** When the mean value of the square of the noise is the same as that of the background, the noise tends to be most visible against a mid-gray background [8]. The mid-gray regions have lower JND's as compared to the other regions.

2.1 Content based classification

2.1.1 Texture and edge analysis.

In a block with abrupt changes between adjacent pixels, the signal energy tends to be concentrated in the AC coefficients. In a flat featureless region of an image the energy is concentrated in the low frequency components. Thus the energy in the AC coefficients can be used as a measure of roughness. Denoting the i,j DCT coefficient as X_{ij} and using the monotonic function log for range compression we arrive at the following expression for the energy in the AC coefficients.

$$E_{AC} = \log \sum_{i,j} (X_{ij})^2 - (X_{00})^2 \tag{1}$$

where X_{00} is the DC DCT coefficient. The maximum energy is in the AC coefficient for a block with a checkerboard pattern with the adjacent pixels having the maximum and minimum permissible gray value. We denote this as E_{max} and is evaluated as

$$E_{max} = \log(G/2)^2 \tag{2}$$

where G is the maximum permissible gray value.

Using E_{max} for the normalization factor we arrive at the measure for the roughness level R for the block b.

$$R_b = \frac{E_{AC}}{E_{max}} \tag{3}$$

The range of R_b is uniformly split into 8 sub groups and the block is given a block distortion index I_b where $I_b \in 1 \ldots 8$ and a high value of I_b implies lower distortion tolerance and vice versa.

2.1.2 Edge Separation and reclassification.

A block that has a large amount of energy in its AC coefficients could either contain prominent edges or be highly textured [12]. The spread of the fluctuations of energy in the block of pixel values would indicate whether it contains an edge or is highly textured. The measure of fluctuations at any location x in the pixel block is given by its gradient, ∇x. In a highly textured block, ∇x would be large at a large number of locations while in a block with prominent edges ∇x would have large values at much fewer locations. By comparing the magnitude of fluctuation to suitably selected threshold λ, it is possible to identify locations of large fluctuations within the image block. A count of these fluctuations can then used to decide if the block has an edge or is highly textured. The threshold count C, is arrived at statistically to make the decision. Therefore a block with large energy or in other words all the blocks satisfying the constraint $I_b \in \{6, 7, 8\}$ have to be subjected to the above explained test and distortion index is reallotted to such blocks with strong edges in the range 1 to 3 based on the edge strength. A block with large R_b has a strong edge if

$$||\{x | x \text{ is a pixel and } mag(\nabla x) > \lambda\}|| < C \tag{4}$$

where $|| \cdot ||$ is the cardinality of the set.

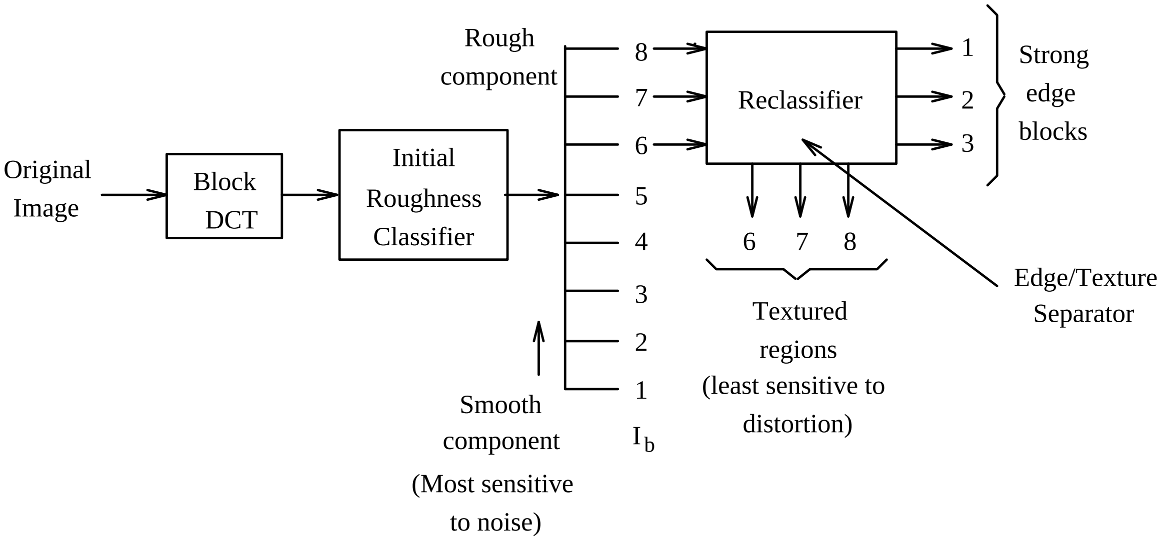


Figure 1: Block diagram for classifying the image sub-block based on noise-distortion sensitivity

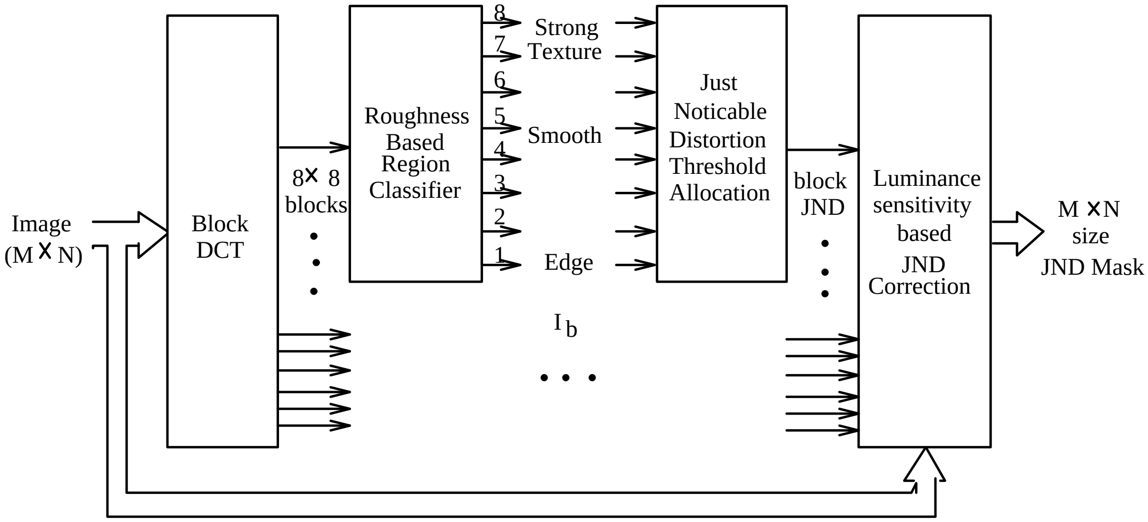


Figure 2: Block diagram for creating a mask of just noticeable distortion values at every pixel location.

Each block based on its corresponding I_b is mapped to a JND value empirically.

$$\hat{J}(b) = K_{I_b} \tag{5}$$

where K_{I_b} varies from 3 to 12 for I_b varying from $1 \ldots 8$ and $\hat{J}(b)$ is the JND value for all the pixels in the block b.

2.1.3 Luminance Sensitivity

The effect of luminance is incorporated into the JND value of the block depending on the gray value of the pixel. Since the distortion in an image is most noticeable in the mid-gray region and sensitivity falls of parabolically as the gray value drifts on both sides we make a final correction to the block JND to give the final just noticeable distortion level at any pixel position i, j in block b.

$$\begin{aligned} J(i,j,b) &= \hat{J}(b) + L(i,j,b) & (6)\\ \text{where } L(i,j,b) &= \lfloor (128 - I(i,j,b))^2/\alpha \rfloor & (7) \end{aligned}$$

α is a predetermined constant and $I(i,j,b)$ is the image pixel gray value at position i, j in block b.

The implementation of the above classification algorithm is illustrated in fig 1 and fig 2.

3 Content based adaptive watermark

We now propose a new method of watermarking of images that exploits both the statistical redundancies and the subjective redundancies in DCT coded images as a modification to the spread spectrum like watermarking technique proposed by Hartung et al [4]. The basic scheme of watermarking is as depicted in fig 3. A noise-like signal is added to the pixel values that is below the threshold of perception decided by our classification algorithm described in the previous section. To embed the information bits $a_j \in \{1, -1\}$ the bits are first spread by a large spread factor cr , called the chiprate. For spreading the information, the bit pattern is repeated in a raster-scan order to tile the entire image.

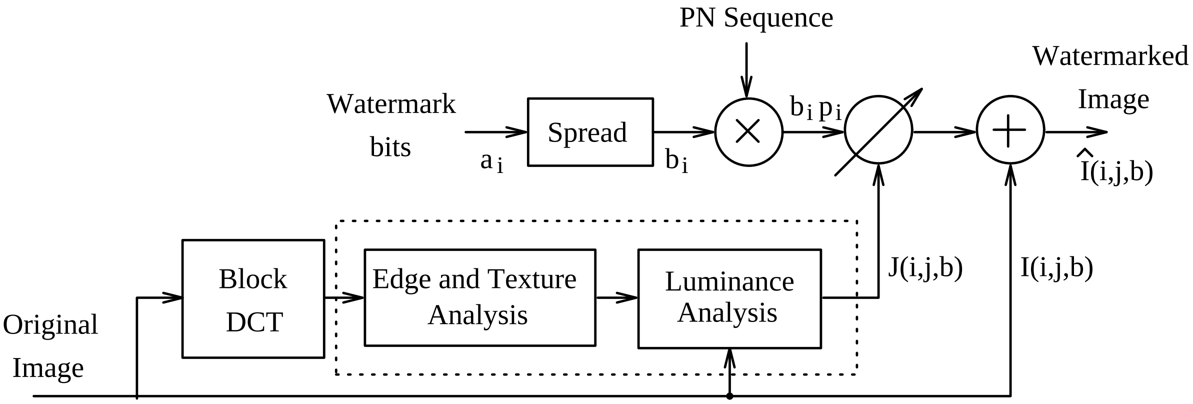


Figure 3: Block diagram of content based perceptual watermarking

This improves its robustness to geometrical attacks such as cropping. The spreading provides for spatial redundancy by embedding the information bits into cr number of pixels.

$$b_i = a_j \quad \forall i = j \times K \tag{8}$$

and K varying from 1 to cr. The spread bits b_i are then modulated with a pseudo-random-noise (PN) sequence.

$$p_i \text{ where } p_i \in \{-1, 1\} \tag{9}$$

This forms the basic watermark sequence. This bit pattern is then amplified by the calculated JND value $J(i,j,b)$ as given in equation 6, corresponding to the raster-scanned image sequence value to which it is to be embedded. The modulated signal, i. e. the watermark $w_i = J(i,j,b) \cdot b_i \cdot p_i$ is added to luminance channel of the raster scanned image I_i yielding the watermarked Image. This ensures that our technique can work for grayscale and color images.

$$\hat{I}_i = I_i + J(i,j,b) \cdot b_i \cdot p_i \tag{10}$$

Since a pseudo-noise sequence is used for modulation the watermark sequence is also noise like and the amplification by HVS dependent value ensures that the watermark is as robust as possible without compromising on the image quality. This also ensures that the watermark is difficult to detect, locate and manipulate.

The detection of the hidden information is done by first buffering the watermarked image. One buffer is maintained for each embedded watermark bit. These buffers are filled with the raster scanned watermarked image pixel values in a cyclic order i.e. the first pixel value is put into the first buffer, the second in the second, the $(N+1)_w$ pixel for a N_w bit length watermark in the first buffer and so on. Once the entire image is scanned each buffer is analyzed by correlating the watermarked image with the same pseudo-noise sequence that was used in the coder (see fig 4) where correlation can be understood as demodulation followed by summation over the correlation window. The correlation window for each bit is the chiprate. If the peak of correlation is positive, the corresponding bit is $+1$ else -1. Considering one buffer with stored watermarked pixel values $\hat{I}_i$ where $i \in 1 \ldots cr$

$$s_j = \sum_{i=1}^{cr} p_i \cdot \hat{I}_i = \sum_{i=1}^{cr} p_i \cdot I_i + \sum_{i=1}^{cr} p_i^2 \cdot J(i,j,b) \cdot b_i \tag{11}$$

The first term on the right-hand side of equation 11 vanishes if

$$\sum_{i=1}^{cr} p_i = 0 \tag{12}$$

i. e. the pseudo-noise sequence contains as many -1's as 1's in the interval $[1 \ldots cr]$, p_i and I_i are uncorrelated. In practice however, the sum in (12) is not zero as the PN sequence and the image pixels are not totally uncorrelated. But choosing a large cr we have adequate redundancy and the summation can be approximated as :

$$s_j = \sum_{i=1}^{cr} p_i \cdot \hat{I}_i + \Delta \approx cr \cdot J(i,j,b) \cdot a_j \tag{13}$$

Δ being the error term. The required information bit $\hat{a_j}$ is

$$\hat{a_j} = sign(s_j) \tag{14}$$

To retrieve the watermark the original image is not necessary but the same unshifted pseudo-noise sequence that was used at the transmitter is required.

4 Implementation and Results

We have implemented the content based watermarking technique as a C program interfaced with the Independent JPEG group's JPEG decoder. It works interchangeably both in the spatial and transform domain. The quality of the watermarked image appears unaltered as in fig 7 and fig 8. The results of the content based image segmentation are shown in fig 5 and fig 6. The fog image was selected as there are areas of fog that are very sensitive to distortion and areas of grass that are less sensitive. As is evident in the figure fig 5 the classification algorithm has segmented the image based on the content's noise sensitivity. The watermark is robust to most image processing operations. Our results show that cropping of the image till 5% had no effect on the recovery of the watermark [12] – hence it is quite robust. The watermark was also recovered without any errors from a JPEG compressed image, compressed to a quality factor 5 (fig 9). The embedded bits were also recovered when the image was subjected to both additive and multiplicative noise, fig 10. We have done extensive experiments which indicate that the perceptual quality of the watermarked image improves with the use of our technique and also the robustness of the watermark improves [12]. Detailed results of our experiments can be found in [12] or at:

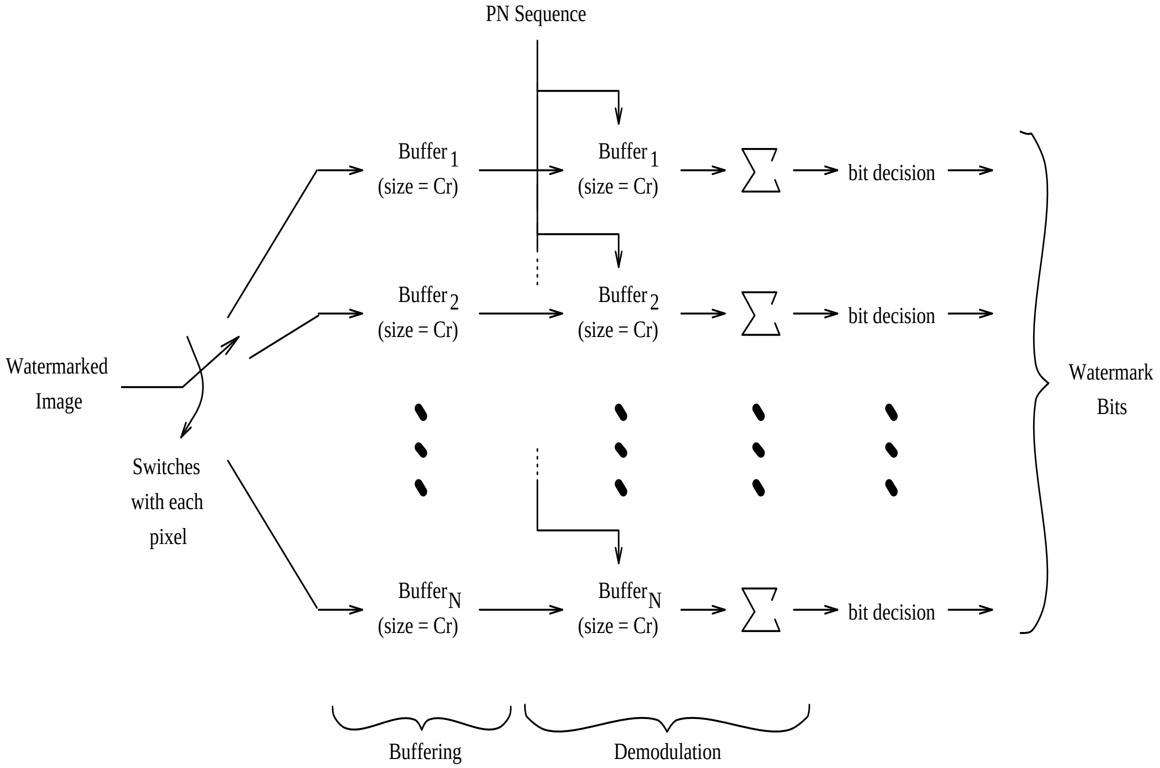


Figure 4: Detection of the embedded 'N' bit watermark

http://minchu.ee.iisc.ernet.in/people/students/rajmohan/.

The high quality images of fog and yose used in our experiments are from:

http://www.kodak.com/digitalImaging/samples.

5 Conclusion

We have proposed a novel and robust method of adaptively watermarking images based on the human visual system (HVS). A new method of classifying a region of the image based on its sensitivity or tolerance to noise has been devised. The underlying content of the image is analyzed to evaluate the noise sensitivity of different regions of the image. The effects of image content features such as texture, edge and luminance have been investigated. Our technique allows the creation of a mask of the size of the image which contains the just noticeable distortion (JND) value for every pixel. In the proposed technique the watermark is spatially spread by a large spread factor to enhance its robustness as is done in spread-spectrum communications. This is then embedded into the image keeping its strength below the mask pixel value. This ensures that the watermark least distorts the regions that are sensitive to changes and exploits perceptual redundancies in the areas of high detail to embed more bits there. Our experiments have shown that the watermark is robust to a wide spectrum of attacks. We feel that such a model will enhance the quality of spatial domain watermarking methods as they currently do not use the HVS models. The advantage of our method is that it can be applied to transform domain techniques also. We are currently working on extending these ideas to MPEG video and are also trying to incorporate other HVS factors, such as contrast masking, in the proposed JND model.

References

[1] J Biemond G C Langellar and J C A Van Der Lubbe. Copy protection for multimedia data based on labeling techniques. 17^{th} *Symposium on Information Theory in the Benelux, Enschede, The Netherlands*, May 1996.

[2] S Craver, N Memon, B L Yeo and M Yeung. Can invisible watermarks resolve rightful ownerships ? *IBM Research Report, RC 20509*, July 1996.

[3] Digimarc Corporation. http://www.digimarc.com

[4] F Hartung and B Girod. Copyright protection in video delivery networks by watermarking of pre-compressed video. *Proc. Multimedia Applications Services and Techniques - ECMAST*, 1997.

[5] T Leighton, I J Cox, J Killian and T Shamoon. Secure spread spectrum watermark for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1686, December 1997.

[6] W J Dowling J J K OR̃uanaidh and F M Boland. Watermarking digital images for copyright protection. *IEE PROC Vision Image Signal Processing*, 143(4), August 1995.

[7] F Jordan, M Kutter and F Bossen. Digital signature of color images using amplitude modulation. *Proc. of SPIE Storage and Retrieval for Images and Video Databases, San Jose, USA*, 3022(5):518–526, Feb 1997.

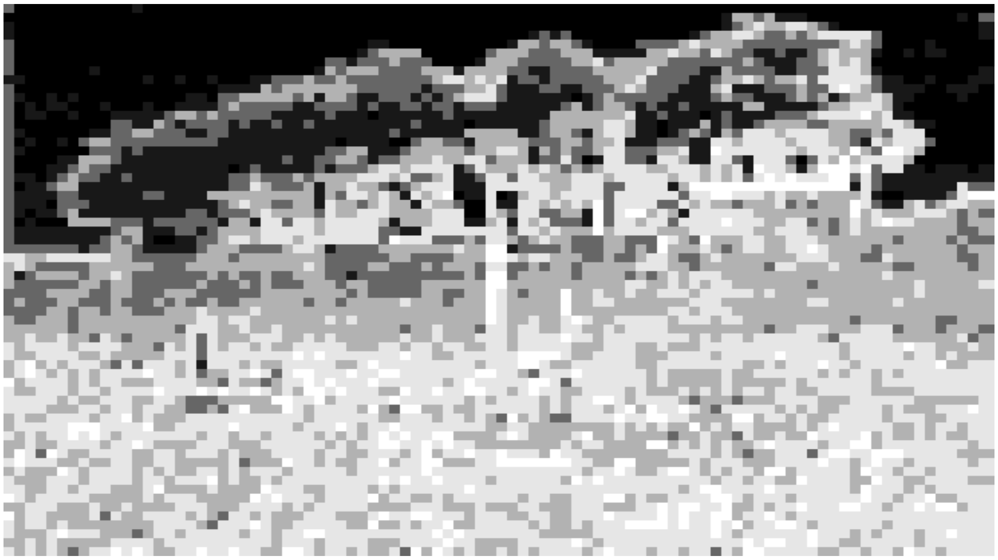


Figure 5: Content based classification of the Fog Image. Highly textured regions are shown in brighter shades of gray and the lesser textured areas are depicted in darker shades of gray. The brightness of the region is inversely proportional to the noise sensitivity of the region.

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

[9] F A P Petitcolas, R J Anderson and M G Kuhn. Attacks on copyright marking systems *Second Workshop on Information Hiding*, Portland, Oregon, April 1998.

[10] C I Podilchuk and W Zeng. Perceptual watermarking of still images. *IEEE Signal Processing Society 1997 Workshop On Multimedia Signal Processing*, 1997.

[11] R D Powell and M J Nitzberg Method and system for digital image signatures. *United States Patent No. 5,721,788.* February 1998.

[12] Rajmohan. Watermarking of digital images. *ME Thesis report, Department of Electrical Engineering, Indian Institute of Science, Bangalore 560012, India*, 1998.

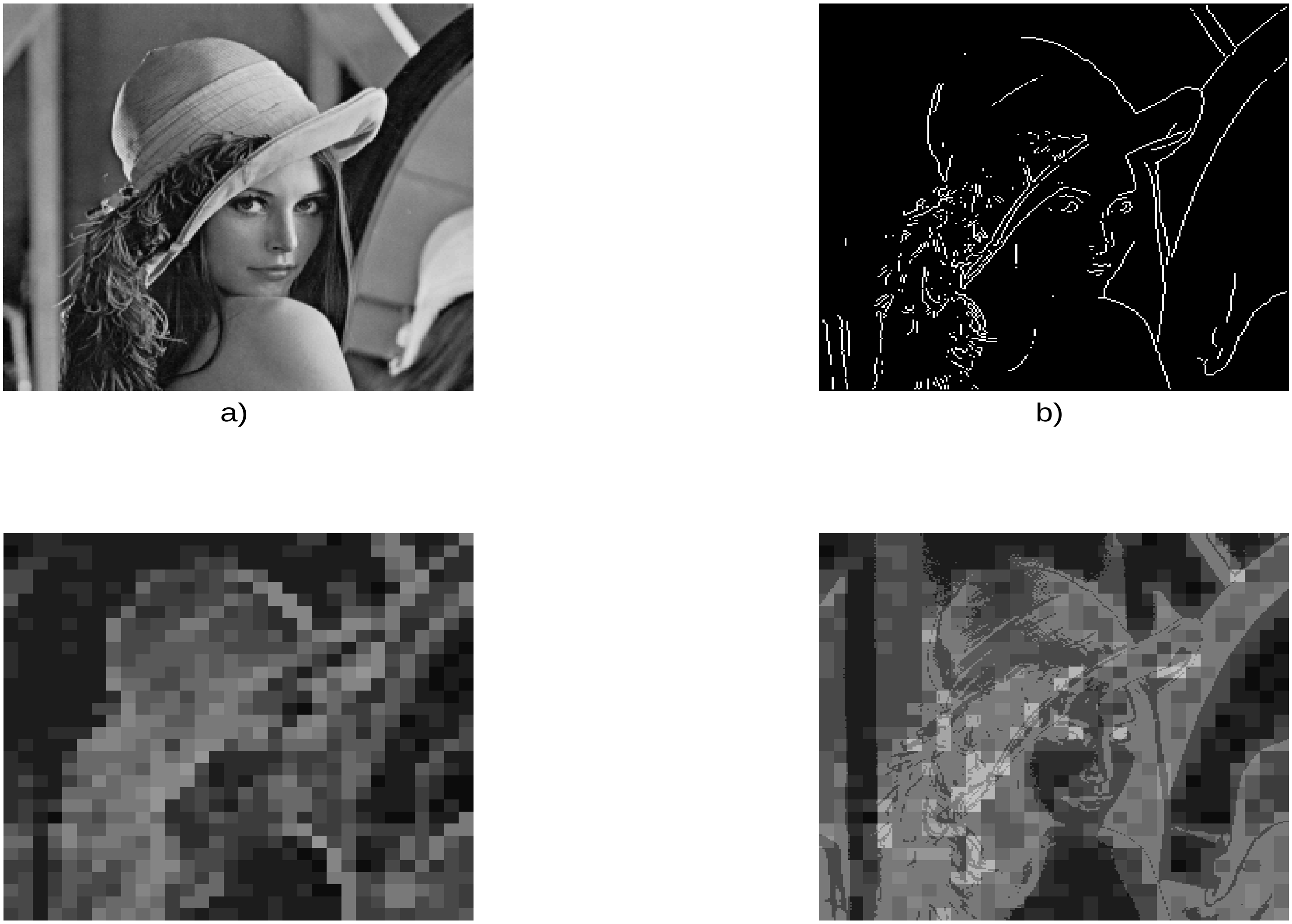
[13] W Raymond and E J Delp. A watermarking technique for digital imagery: Further studies. *http://dynamo.ecn.purdue.edu/ ac/delp-pub.html.*

[14] B Tao and B Dickenson. Adaptive watermarking in the dct domain. *International Conference on Acoustics and Signal Processing*, 1997.

[15] L Torres and M Kunt. *Video Coding, The Second Generation Approach.* Kluwer Academic Publishers, 1995.

[16] G Wallace. The joint photographic experts group – still picture compression standard. *IEEE Transactions on Consumer Electronics*, 38(1), 1992.

[17] A B Watson. Dct quantization matrices visually optimized for individual images. *Proceedings Of SPIE Conference on Human Vision, Visual Processing and Digital Display IV ,San Jose, 1992*, 1992.



a)

b)

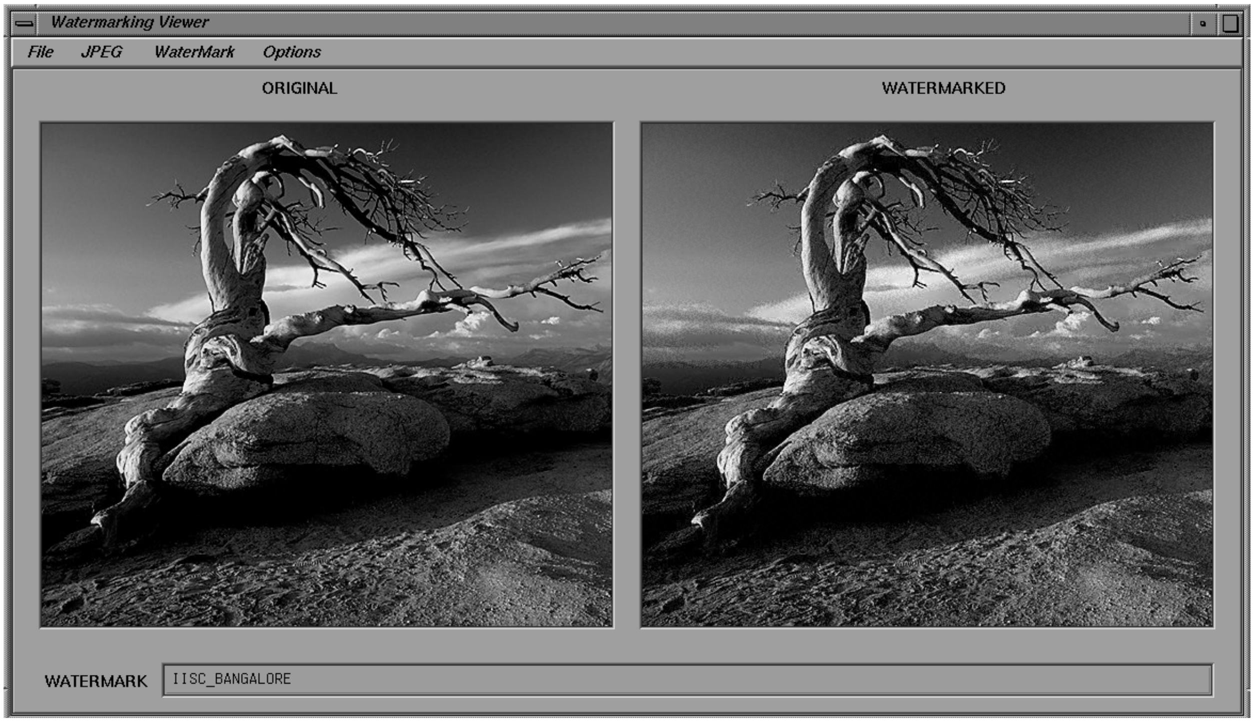
c)

d)

Figure 6: Perceptual region classification of lena image. a) Original image b) Prominent edges of the image. c) Classification based on the texture as given by the energy of the DCT AC coefficients. d) Classification based on texture, edge and gray level of the pixel. Bright areas are less sensitive to noise and vice-versa.

Key #2 2259

Figure 7: Perceptually watermarked lena image with average power of 6 per pixel.



WATERMARK IISC_BANGALORE

Figure 8: Content based invisible watermark.

Figure 9: Effect of JPEG coding of the watermarked image on the detection of the watermark. JPEG coded to 5 % quality factor and watermark successfully extracted.

Figure 10: Addition of gaussian noise (0,15) to the watermarked image compressed to quality factor of 50. Watermark was extracted with no errors.