Robust Digital Image Watermarking using Region Adaptive Embedding Technique

Chunlin Song, Sud Sudirman and Madjid Merabti
School of Computing and Mathematics Science
Liverpool John Moores University
Liverpool, UK
c.l.song@2004.ljmu.ac.uk, {s.sudirman, m.merabti}@ljmu.ac.uk

Abstract: **Improving the robustness of watermark in withstanding attacks has been one of the main research objectives in digital image watermarking. In this paper we propose a novel region-adaptive watermarking technique that can provide improvements in both robustness and visual quality of the watermarks when compared to the original, non-region-adaptive, embedding technique. The proposed technique, which is derived from our previously published research finding, shows that the relative difference in spectral distributions between the watermark data and the host image plays an important role in improving the watermark robustness and transparency.**

***Keywords* – Quad-tree algorithm, discrete wavelet transform, region-based watermarking system.**

I. Introduction

As one of the most popular and viable techniques in protecting copyrights in digital media, watermarking technology has received enormous level of attention of researchers and practitioners alike. Unfortunately, due to the same reason, watermarking technology has also attracted the attentions of hackers and criminals alike who are interested in breaking the watermarks in order to crack the copyright protection system. As a result, there is a constant challenge on the researchers to keep improving the robustness of the watermarking technique while at the same time maintaining its transparency as to not intruding any legitimate use of the media.

Progress in this area has been steady as can be seen from a healthy number of publications in the field and the sheer number of institutes around the world that deal with the issue [1]. In the more specific field of digital image watermarking, one of the most notable techniques is region-based image watermarking [2]. The paper described a method for embedding and detecting chaotic watermarks in large images. An adaptive clustering technique is employed in order to derive a robust region representation of the original image. The robust regions are approximated by ellipsoids, whose bounding rectangles are chosen as the embedding area for the watermark. The drawback of this technique is due to limited number of suitable regions for storing the watermark the watermark storing capacity can be low.

In this paper, we present a novel watermarking technique which works by adaptively embedding the watermark data into different region of the host image. The rationale of our approach is based on the research finding we came into in our previous work [3][4]. This finding will be described in detailed in this paper for convenience.

II. Rationale of the Region-Adaptive Approach

Our previous work in this field dealt with determining the effect of different removal attacks on watermark signals. The experiment conducted involved subjecting watermarked images to a number of watermark removal attacks and observe the results in both spatial and frequency domains.

In that work, we analyzed the effect of seven different watermark attacks namely Gaussian noise attack, salt and pepper noise attack, Gaussian smoothing attack, sharpen attack, histogram equalization attack, median filter attack and JPEG compression attack.

The analysis was carried out using two image analysis tools namely image histogram and Fourier Transforms. In the image processing context, the histogram of an image refers to the distribution of the pixel intensity values. This histogram is a graph showing the number of pixels in an image at each different intensity value found in that image. Fourier transform converts image data into its frequency spectrum. As with many other signal transforms, Fourier Transform converts time/spatial domain signals into alternative representations that are more amenable for certain types of analysis. The Fourier spectrum of an image is often visualized as a grayscale image whose intensity corresponds to the strength or magnitude of spectrum. The coordinate of these pixels corresponds to the frequency in x-y directions. The low frequency is located at the centre of the image and high frequency is located around the edge. Fourier spectrum has an important characteristic in which its total energy is preserved. This means both low and high frequency components are complementary to each other. In other words, if the high frequency component of an image increases, the low frequency component of that image subsequently decreases. Therefore, any shift in frequency band can be easily detected as the change in spectrum strength in both low and high frequency regions.

The analysis of removal attacks on watermarks is summarized in Table I. Our analysis indicates that the watermark attacks could be divided into two general categories: high frequency (HF) watermark attack and low frequency (LF) watermark attack. The former category

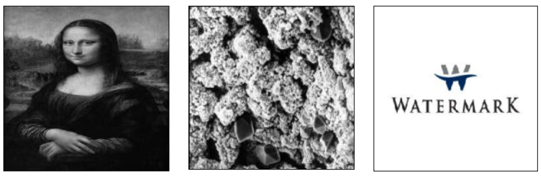
ISBN: 978-1-902560-25-0 © 2011 PGNet

attacks the high frequency part of the signal while keeping most of the low frequency component relatively unchanged and the latter category attacks vice versa. Gaussian noise attack, salt and pepper noise attack, median filter attack, sharpen attack and JPEG compression attack belongs to LF watermark attack, whereas Gaussian smoothing attack and histogram equalization attack are more akin to the HF watermark attack.

TABLE I. EFFECTS OF DIFFERENT REMOVAL WATERMARK ATTACKS IN SPATIAL AND FREQUENCY DOMAINS [3]

Watermark attacks	Effect in spatial domain	Effect in frequency domain
Gaussian smoothing attack	Reduces the variation in image pixel values.	Acts as low pass filter
Gaussian noise attack	Increases the variation in pixel values	Similar effect to a high pass filter
Salt & pepper noise attack	Same as Gaussian noise attack	Same as Gaussian noise attack.
Median filter attack	Similar, albeit much smaller, effect as with Gaussian smoothing attack.	Similar effect to a high pass filter
Histogram equalization attack	Reduces the number of unique grayscale values and make the histogram more uniformly distributed	Similar, albeit more moderate, effect as Gaussian smoothing attack
Sharpen attack	Reduces the overall image intensity and amplifies differences around edges	Acts as high pass filter
JPEG Compression attack	Reduces the variation in image pixel values and creating blocks, or uniform regions, in the image.	Similar effect to a high pass filter.

This finding suggests that the spectral distribution of the watermarked image plays an important role in the overall robustness of the watermark algorithm to a particular attack. Based upon this finding, we hypothesized that in order to maximize the robustness of the watermarked data, the spectral distribution of the host data and the watermark data should be similar. Furthermore, in order to counter both types of attacks the watermark data should both possess strong high frequency and low frequency parts.



(a) (b) (c)

Figure 1. (a) The host image, (b) the HF watermark image and (c) the LF watermark image

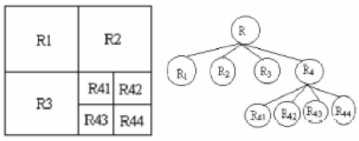
To test this hypothesis we develop our proposed region-adaptive watermark technique. This technique employs two unique solutions namely a) it uses two watermark images, each with strong high frequency or low frequency components and b) it embeds different parts of the watermark images into the host image based on the difference of their spectral distributions. Examples of the host image and High Frequency (HF) and Low Frequency (LF) watermark images used in our experiment are shown in Figure 1. The details of our region-adaptive watermarking technique are described in the next section.

III. REGION-ADAPTIVE WATERMARKING

A. *Watermark Insertion*

The region-adaptive watermarking technique we propose in this paper works by embedding parts of the watermark image into selected regions in the host image. The selection process works by matching the watermark and host image regions with similar spectral distributions. To improve the watermark embedding and extraction speed, we decide not to use regions with arbitrary shape and orientation. Instead, we will use non-overlapping squares of varying sizes as our regions.

To do this we first divide the host image using Quad Tree Partition technique. Assuming that the host image is a square image, Quad Tree Partition works by dividing the image into four equal sized square blocks. A test is carried out on each block to check whether it meets the criterion of homogeneity. If a block meets the criterion, it will not be divided further otherwise it will be subdivided again into four blocks. This process is repeated until every block meets the criterion. The result may have blocks of several different sizes. Figure 2 illustrates a quad tree partition of an image up to 2 levels. The parent node in the tree represents the entire image region and its four descendant's nodes represent the large disjoint sub regions.



(a) (b)

Figure 2. Quad Tree Partition (a) Resulting Blocks, (b) The quad tree structure

The criterion used in deciding the partitioning of a region takes into account the pixel characteristics of that region. In general, the test checks if the characteristics of pixels in that region are too heterogeneous then the test will yield negative and vice versa. To achieve this, an image segmentation algorithm is applied to the host image prior to the partitioning process. Each segment of the host image should represent areas in which pixel characteristics are homogenous. By performing this process, we also avoid computationally expensive task of recalculating pixel characteristics in each region during the partitioning process.

There are a number of image segmentation algorithms that can be used such as Blobworld [6], Gabor Filters [7], Markov Random Fields [8] to name a few. In this paper, the Markov Random Field (MRF) technique as described in [8] is used due to its computation speed and statistical categorization of image textures.

The entire watermark embedding process is described below:

1. Segmentation of host image using MRF technique.
2. Quad tree partition of the segmented host image.
3. Partition the watermark images into regions of similar size as the partition in (2).
4. Find the best match between the host image partition and the watermark image partition.
5. Embed the watermark region into its matching host image region.
6. Store the location and size of the watermark insertion region. This can be stored in the header of the watermark image file such as EXIF in JPEG.

To embed the watermark we employ our modified version of the discrete wavelet transform (DWT) watermarking algorithm [5]. The host image is first decomposed into its 4^{th} level wavelet and approximation coefficients using DWT. The watermark consisting of pixel values inside regions of the watermark image is inserted by modifying the wavelet coefficients belonging to the three detail bands. Since quad tree is used to partition the host image, the wavelet coefficients of the resulting partition can be easily extracted from the wavelet coefficients of the entire host image.

B. *Watermark Extraction*

The watermark extraction of the proposed region-adaptive technique is less complex than the insertion process. The extraction process does not require the watermarked image to be segmented and partitioned. However, it requires the information on where and how big the watermark regions are. The complete process is listed below:

1. Extract the stored location and size of watermarked regions.
2. Extract the watermark data out from the watermarked blocks.
3. Merge the extracted watermark data to create the complete watermark image.

IV. Experimental Results and Analysis

We applied the region-adaptive watermark technique described in the previous section to the images shown in Figure 1. The host image has dimension of 512x512 pixels and both watermark images have dimension of 256x256 pixels. The HF watermark image is divided into 64 blocks of 32x32 pixels and the LF watermark image is divided into 16 blocks of 64x64 pixels.

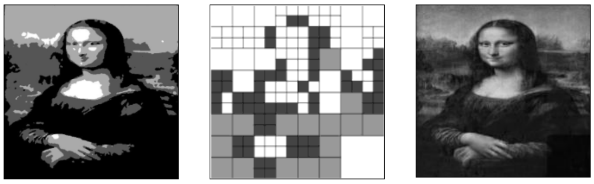
As a quantitative measure of the degradation effect caused by the attacks we use Peak-Signal-to-Noise Ratio (PSNR). The formulation between the original and the attacked watermarked signals can be found described in [3-4]. High PSNR values indicate lower degradation hence indicating that the watermarking technique is more robust to that type of attack.

Two experiments were conducted to test the algorithm. The first experiment is aimed to verify that inserted watermarks images can be extracted with minimal distortion. The second experiment measures the robustness of the proposed region-adaptive technique and compares it to the original DWT algorithm [4].

A. *Watermark insertion and extraction verification.*

The first experiment try to prove that watermark images could be inserted and extracted completely by using proposed region-adaptive technique. Figure 3 shows some intermediate results of the watermark insertion process. The figure shows the segmented host image, the parts of the host image where the different watermarks are inserted and the resulting watermarked image. Figure 4 shows the extracted watermark images.

As can be seen from Figure 4, the extracted watermark images are similar to the original watermark images shown in Figure 1.

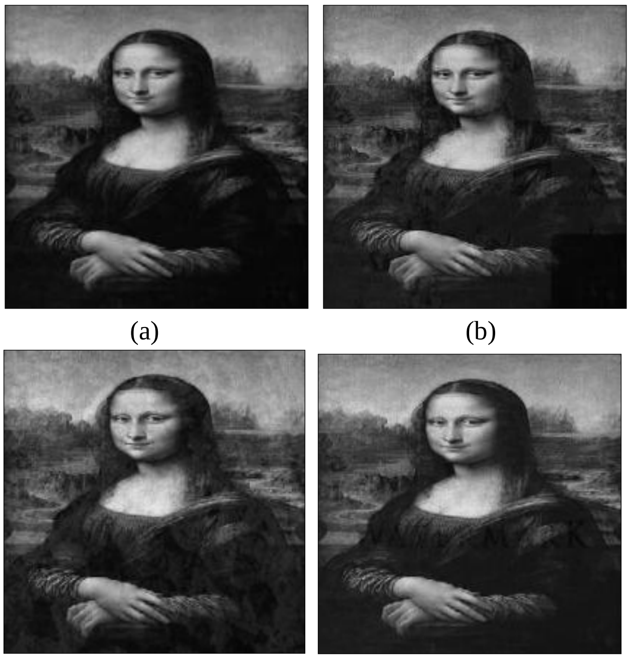


(a) (b) (c)

Figure 3. (a) MRF segementated host image, (b) watermark insertion region (dark grey – HF and light grey – LF) and (c) watermarked image

(a) (b)

Figure 4. Extracted (a) HF watermark, and (b) LF watermark images



(a) (b)

(c) (d)

Figure 5. Original watermarked image by different algorithm: (a) region-based algorithm. (b)HF watermarked image (c) LF watermarked image

To show the quality of the watermarked output we show in Figure 4 the different watermarked image created using the proposed region-adaptive technique and those created using the original DWT technique. The Figure shows that there are fewer artifacts introduced, especially around the low frequency region of the watermarked image, using the region-adaptive technique compared to those using the DWT technique. This result can also be quantitatively measured by the higher PSNR values of the proposed technique as shown in Table II.

TABLE II. PSNR VALUES OF UNMODIFIED WATERMARKED IMAGE

Region based	**DWT (HF watermarked)**	**DWT (LF watermarked)**
27.44	24.06	25.33

B. *Robustness comparison*

To test the robustness of the proposed watermarking scheme, seven watermark removal attacks are applied to the watermarked image. They are Gaussian noise, salt and pepper noise, sharpen, smoothing, median filter, histogram equalization and JPEG compression attack. The severity of these attacks can be adjusted by modifying their corresponding parameter values. Definitions of these parameters can be found is given in [3]. Figure 6 and 7 show the extracted watermark images after different attacks.

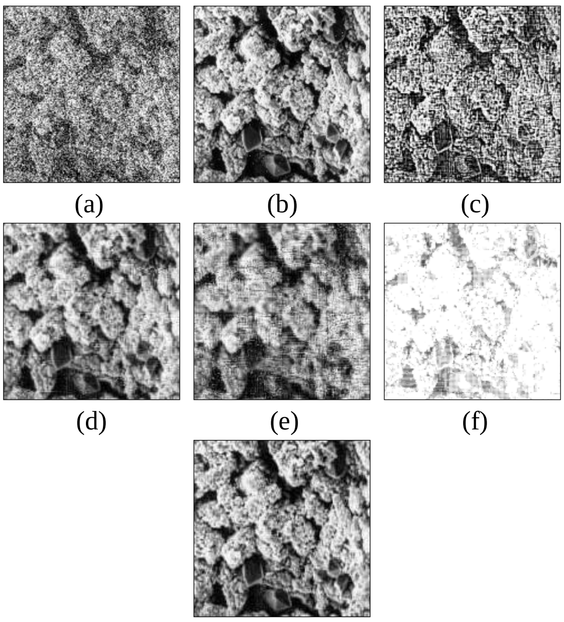


Figure 6. Extracted HF watermark image (a) Gaussian noise with σ = 0.005 (b) Salt and Pepper Noise with σ= 0.005 (c) Sharpen with σ = 1.0 (d) Smoothing with σ = 200 (e) Meidan filter (f) Histogram Equalization with η= 200 (g) JPEG Compression with σ = 90

To compare the robustness of the proposed technique with the original DWT technique, we conducted the experiment using 50 different host images. The parameters of the attack algorithms used in the experiment are the same as those used in Figure 6 and 7. The PSNR values between the unmodified watermark image and the attacked watermarked image are then averaged. These results are summarized in Table III.

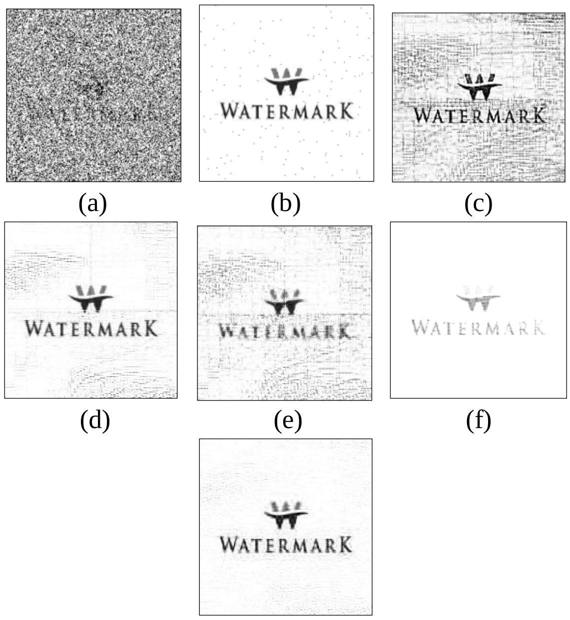


Figure 7. Extracted LF watermark image (a) Gaussian noise with σ = 0.005 (b) Salt and Pepper Noise with σ= 0.005 (c) Sharpen with σ = 1.0 (d) Smoothing with σ = 200 (e) Meidan filter (f) Histogram Equalization with η= 200 (g) JPEG Compression with σ = 90

TABLE III. PSNR VALUES OF THE ATTACKED WATERMARKED IMAGE

Attack Method	**Region based**		**DWT**	
	HF	*LF*	*HF*	*LF*
Gaussian noise	7.89	5.06	7.84	5.09
Salt & pepper noise	27.8	26.0	27.6	26.2
Sharpen	8.27	6.61	7.71	5.17
Smoothing	14.9	14.9	22.0	18.8
Median	10.69	10.41	10.71	8.39
Histogram Equalization	6.91	8.68	6.86	6.90
JPEG Compression	19.7	22.1	18.6	15.6

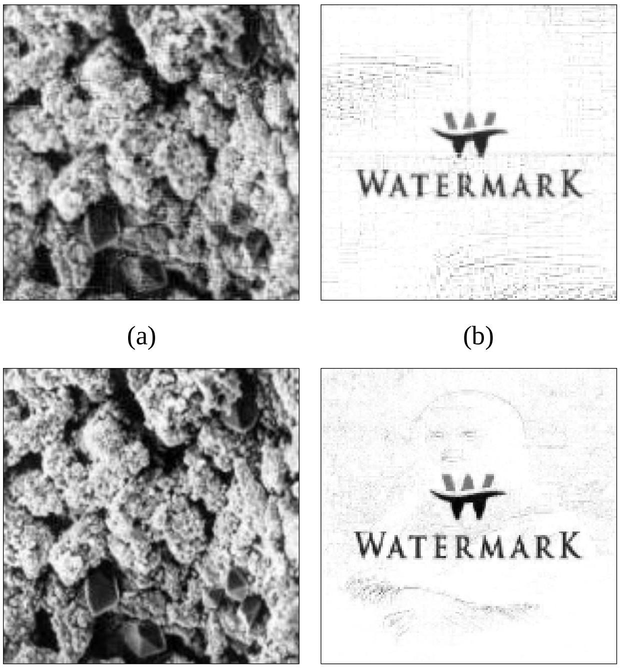
C. *Analysis of results*

As can be seen from the results shown in Table III, the proposed region-based approach produced higher or equal PSNR values in most types of attacks. The results also show that the proposed technique fares worse in one specific type of attack namely the Gaussian smoothing attack. The summary of the comparison is given in Table IV.

Although it appears that the proposed technique fares worse when subjected to the Gaussian smoothing attack than the DWT technique, we would like to note a couple of arguments which support our claim that our technique is still overall a better watermarking technique.

TABLE IV. THE ROBUSTNESS LEVEL OF THE PROPOSED TECHNIQUE AS COMPARED WITH THE ORIGINAL DWT TECHNIQUE

Better	Similar	Worse
HF & LF in sharpen	HF & LF in Gaussian noise	HF & LF in smoothing
HF & LF in histogram equalization	HF and LF in salt and pepper noise	
HF & LF in JPEG compression	HF in median filter	
LF in median filter		



(a) (b)

(c) (d)

Figure 8. Extracted HF & LF watermark image after smoothing attack in region-based algorithm (a)(b) and DWT (c)(d)

Firstly, we would like to point out that our technique inserts twice as much watermark data to the host image as the DWT technique. This is because our technique inserts both HF and LF watermark image at the same time whereas the DWT technique stores one image at a time.

Furthermore, we can show that the extracted watermark images stored using our technique were less distorted than those using the DWT technique. An example of our claim can be seen in Figure 8. The Figure shows that the smoothing attack imprints some significant artifacts on the extracted watermark image when it is inserted using the DWT technique. This can be clearly seen as the face contour in the extracted LF image.

V. CONCLUSION

We have presented in this paper a novel digital image watermarking technique using region adaptive approach. The technique is derived from our previous work [3][4]. Our hypotheses are:

1. By making the spectral distribution of the host data and the inserted watermark data similar we will be able to maximize the robustness of the watermarked data.
2. In order to counter both high frequency and low frequency type attacks the watermark data should both possess strong high frequency and low frequency components.

Our region adaptive watermarking technique is realized by using two watermark images, each with a strong High Frequency or Low Frequency components. Non overlapping regions of these watermark images are inserted into the host image using a combination of image segmentation and Quad-Tree partitioning techniques.

Our experimental results show that the region adaptive technique outperforms the original DWT technique [5] in withstanding attacks. Furthermore, our technique produces better visual quality watermarked image for relatively similar distortion level. These are in addition to the fact that our technique can store twice as much watermark data than the latter technique.

Further improvement to the region based watermarking technique is planned, this includes improving its robustness against geometric attacks such as rotation, translation and scaling.

REFERENCES

[1] Petitcolas, F.A.P., Anderson, R.J., Kuhn, M.G., Information Hiding—A survey, Proceeding of the IEEE, Special Issue on Protection of Multimedia Content, 1062-1078, July 1999.

[2] A. Nikolaidis and I. Pitas, "Region-based image watermarking," *Image Processing, IEEE Transactions on*, vol. 10, no. 11, pp. 1726-1740, 2001

[3] Cl.Song, S.Sudirman and M.Merabti, "A Spatial and Frequency Domain Analysis of the Effect of Removal Attacks on Digital Image Watermarks", Proc 11th of PostGraduate Network Symposium, 119-124, June, 2010.

[4] C.Song, S. Sudirman, M.Merabti and D.L.Jones, "Analysis of Digital Image Watermark Attacks", 6th IEEE International Workshop on Digital Rights Management, 2010.

[5] P.N.Tao and Eskicioglu, "A robust multiple watermarking scheme in the Discrete Wavelet Transform domain", Internet Multimedia Management System Conference, vol. 5601,133-144, 2004.

[6] Chad, C., Magan, T., Serge, B., Joseph H., Jitendra M.: Blobworld: A System for Region-Based Image Indexing and Retrieval, Visuan Information and Inforamtion Systems, Lecture Notes in Computer Science, Vol: 1614/1999, 1999.

[7] Jain, A.K., Farrokhnia, F.: Unsupervised Texture Segmentation Using Gabor Filters, Systems, Man and Cybernetics, Conference Proceedings., 14-19, Nov, 1990

[8] Erdogan, C., Deliang, W,: Texture Segmentation Using Gaussian-Markov Random Field and Neural Oscillator Networks, IEEE Transactions on Neural Networks, Vol. 12, No. 2, 394-404, March 2001.