Digital Image Watermarking based on Regularized Filter

Kharittha Thongkor, Pipat Supasirisun and Thumrongrat Amornraksa
Multimedia Communication Laboratory, Computer Engineering Department, Faculty of Engineering, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
kharittha@gmail.com, pipat@cpe.kmutt.ac.th, t_amornraksa@cpe.kmutt.ac.th

Abstract

This paper presents a spatial domain image watermarking method based on regularized filter. In the proposed method, a watermark image is embedded into a host color image directly by modifying the blue color component. The watermarkmark strength is controlled by two factors, i.e. a constant value and the luminance within a local embedding area. The prediction of the original host image is obtained from the watermarked image by using the regularized filter, so that the embedded watermark can be blindly recovered by substracting the predicted image from the watermarked image. Set of experiment are carried out to verify our proposed method. The results show that the accuracy of the extracted watermark in terms of NC was higher, compared to the previous method. The robustness comparison between both methods against different types of attacks is also performed and presented.

1. Introduction

Most information in form of digital media at present, such as audio, video and images can be rapidly distributed all over global networks. People sometimes misuse such media by distributing the illegal made copies to their friends, and even to the public, leading to a copyright related problem. In order to prevent this kind of problem, digital watermarking is introduced and used widely. Digital watermarking is a technique used to embed secret information called "watermark" into host digital media before distributing it to the public while quality degradation of the watermarked media must be unnoticeable by the human eye. Since the embedded watermark still exist within the new copies of that watermarked media, the watermark inside can then later be recovered, and used to verify the real owner. In practice, some noises and attacks can be introduced unintentionally and/or intentionally. A decent watermarking method should then be robust against all possible noises and attacks.

Nowadays, a large number of image watermarking methods have been proposed and proved to be robust against various kinds of noises and attacks. Basically, the methods can be classified into two main groups: frequency domain and spatial domain based watermarking. In the frequency domain, the watermark is embedded by modifying the coefficients obtained from the transformed image pixels so that the embedded watermark can survive most compression standards, e.g. JPEG and JPEG2000 [1], [2]. However, the demonstration of many researches showed that the frequency domain based approach was not robust enough against geometrical attacks, e.g. cropping, and a few watermark bits can be added to some particular frequency ranges of the host image. In contrary, the processes of watermark embedding and extracting in the spatial domain are simple to perform by directly modifying the host image pixels. Many studies have shown that an embedded watermark can survive most geometrical attacks and a large number of watermark bits can be added to the host image. For example, M. Kutter *et al.* [3] proposed an image watermarking method based on amplitude modulation. Their method was proved to be robust against various types of attacks including JPEG compression standard. The method embedded a watermark bit into an image pixel by modifying the blue color component in that pixel using either additive or subtractive depending on the watermark bit value. Accordingly, the blue color component was selected to carry the watermark bit because it is the one that human eye is least sensitive to. The watermark extraction was blindly achieved by using a prediction technique based on a linear combination of pixel values in a neighborhood around the embedded pixels. The predicted original pixel value was then subtracted from the watermarked one to obtain the embedded watermark. To improve the method, T. Amornraksa *et al.* [4] proposed some techniques to enhance its watermark extraction performance, i.e. by balancing the watermark bits around the embedding pixels, tuning the strength of embedding watermark in according with the nearby luminance, and reducing the bias in the prediction of the original image pixel from the surrounding watermarked image pixels. They also demonstrated how to embed a watermark image (logo) with the same size as the host color image. However, inaccurate prediction of original image pixel in some particular type of images containing high variance image pixel values resulted as a low accurate extracted watermark. C.-H. Lai *et al.* [5] proposed the image watermarking method based on multi-scale neighbourhood matching which solved the problem of inaccurate original image pixel prediction. Their method checked the image pixel value based on the standard deviation of its neighbourhood to determine a suitable image pixel to perform the watermark embedding. With the watermark positions map sent from the embedder to the detector, the watermark extraction was simply achieved by differentiating between the embedded pixel value and the average of its neighbourhood. Although this method achieved a high quality of the extracted watermark, it provided low watermark capacity compared to the number of host image pixels. In addition, apart from requiring the positions map at the watermark detector, the method did not take any benefit from the human visual system. J. A. Hussein [6] proposed an alternative watermarking method based on log-average luminance, whereby 8×8 pixels blocks were used for watermark embedding. These blocks were chosen spirally from the center of the embedding image and had a log-average luminance greater than or equal to the log-average luminance of the entire image. However,

apart from using an inconvenient non-blind approach in this scheme, modifying the luminance components of host image significantly degraded the visual image quality.

In this paper, we carefully study the characteristics of the above methods, and propose a new digital watermarking based on regularized filter for color images. Precisely, the regularized filter is used as a tool to determine the original un-watermarked image. We carry out some experiments to evaluate the accuracy of the extracted watermark from our proposed method, and compared it to the previous work in [5]. The next section describes details of the proposed watermarking method.

2. Proposed Watermarking Method

A watermark bit, *w*, is embedded into a host color image pixel directly by modifying its blue color component value, *B*, using either additive or subtractive depending on *w*, and proportional to the luminance of the embedding pixel, *L*, $B' = B + wL$, where B' is the watermarked component. *L* is used for tuning the strength of *w* due to the fact that changes in high luminance pixels are less perceptible to the human eye, so that more energy of *w* can be embedded to achieve a higher level of robustness. The extracted watermark, w', was blindly achieved by using a prediction technique based on a linear combination of blue color component values in the eight neighborhood around the embedded components as shown in the following equation.

$$w'(i,j) = B'(i,j) - \left[\frac{1}{8} \left(\sum_{m=-1}^{1} \sum_{n=-1}^{1} B'(i+m, j+n) - B'(i,j) \right) \right] \quad (1)$$

The similar techniques proposed in [4] was also applied to enhance the watermarking performance, i.e. by balancing the watermark bits around the embedding components with a random pattern bits, tuning the strength of embedding watermark in according with the luminance value weighted from a Gaussian pixel weighting mark, L_{gauss}, and replacing the neighboring pixel that most differs from $B'(i,j)$ with $B'(i,j)$ in the process of original blue component prediction. The watermark embedding and extraction equations of the proposed method are given as

$$B' = B + swL_{gauss}, \quad (2)$$

$$w'(i,j) = B'(i,j) - \left[\frac{1}{8} \left(\sum_{m=-1}^{1} \sum_{n=-1}^{1} B'(i+m, j+n) - B'(m\_\max, n\_\max) \right) \right], \quad (3)$$

where *s* is a scaling factor used to tune the strength of *w* for the whole image. $B'(m\_\max, n\_\max)$ is a neighboring component around *(i,j)* that most differs from $B'(i,j)$. By considering that embedding *w* into $B(i,j)$ is similar to adding noise to $B(i,j)$, and $B'(i,j)$ can be considered as a noisy image formed by the addition of noise, $n(i,j)$, to $B(i,j)$, i.e. $B'(i,j) \approx B(i,j) + n(i,j)$. The prediction of $B(i,j)$ can then be obtained directly from a proper filtered image. Theatrically, regularized filter assumes that is B' was created by convolving B with a point-spread function (PSF) together with adding noise. Thus, PSF has to be defined before performing the regularized filter. From (2), the embedded watermark consists of Gaussian distribution property, i.e. the effect of Gaussian pixel weighting mark; we define the PSF as

$$PSF = h(i,j) = e^{\frac{-(i^2+j^2)}{2\sigma^2}} \Big/ \sum_i \sum_j e^{\frac{-(i^2+j^2)}{2\sigma^2}}, \quad (4)$$

where *i* and *j* is the distance from the origin in the horizontal and vertical axis, respectively. σ is the standard deviation of the zero mean Gaussian distribution. The regularized filter with such PSF is thus used to remove $n(i,j)$ from $B'(i,j)$, and the output filtered image $B''(i,j)$ is given as.

$$B''(i,j) = h(i,j) * B'(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} h(i,j) B'(i-m, j-n), \quad (5)$$

where $*$ denotes the convolution operation, $h(i,j)$ is the PSF, *M* and *N* are the numbers of row and column in B', respectively. Note that the more knowledge about $h(i,j)$, the closer $B''(i,j)$ will be to $B(i,j)$. *w* can then be extracted by subtracting $B''(i,j)$ from $B'(i,j)$.

To describe our watermark embedding process in a practical system (see Fig. 1), the watermark image, $I_w(i,j)$, is created using only two colors: black (0) and white (1). The watermark bits are first permuted to disperse bits 0s and 1s. Then the numbers of bit 0 and 1 are balanced by XORing the result with a pseudo-random bit stream generated from a key-based stream cipher. Finally, bits 0s are converted into -1, so that the watermark to be embedded become as $w(i,j) \in \{-1,1\}$. The watermark embedding is performed by modifying the image pixel in blue color component of the host image, *B(i,j)* in a line scan fashion, left to right and top to bottom. The modification of *B(i,j)* are either additive or subtractive, depending on three factors, i.e. *w(i,j)*, *s*, and proportional to L_{gauss}.

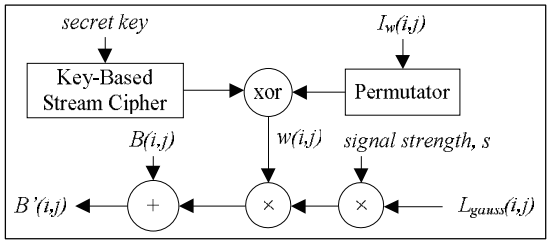

Fig. 1. Block diagram of the proposed embedding process

For the watermark extraction process, the prediction of *B(i,j)* is achieved by using the regularized filter with the window size and standard deviation (σ) of $h(i,j)$ set to be 3 and 0.16, respectively. These values were empirically obtained from a large number of trial-and-error approaches. Then, the estimate of *w(i,j)* is obtained by subtracting $B''(i,j)$ from $B'(i,j)$. Since *w'(i,j)* can be either positive or negative, its sign is used to estimate the value of *w(i,j)*. That is, if *w'(i,j)* is positive (or negative), *w'(i,j)* is estimated as 1 (or -1, respectively). Note that the magnitude of *w'(i,j)* reflects a confident level of estimating *w(i,j)*. Finally, the bits -1s of *w'(i,j)* is converted into 0, and the result is XORed with the same pseudo-random bit stream, as used in the embedding process and then the result is passed to invert permutation to obtain the extracted watermark image, $I'_w(i,j) \in \{0,1\}$.

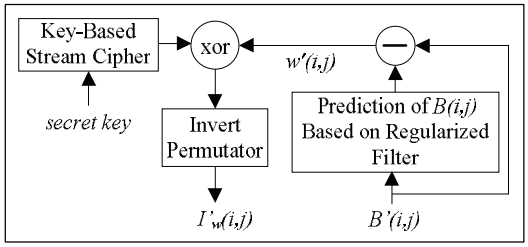

Fig. 2. Block diagram of the proposed extraction process

3. Experimental Setting and Results

In the experiments, ten standard 256×256 pixels color images having different characteristics were used as the host images. '*tiffany*', '*tower*' and '*splash*' represented images with a low level of detail, '*pepper*', '*lena*', '*boats*' and '*girl*' represented images with a medium level of detail, and '*barbara*', '*baboon*' and '*sanfrancisco*' represented images with a high level of detail. A 256×256 pixels binary image containing our university's logo was used as a watermark. A common objective quality measure called Peak Signal-to-Noise Ratio (PSNR) was used for evaluating the quality of watermarked image. After we extracted the watermark, its quality in terms of NC (Normalized Correlation) was measured and used for comparison purpose. The NC formula is given by

$$NC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} I_w(i,j)I'_w(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} I_w(i,j)^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} I'_w(i,j)^2}}, \quad (6)$$

where $I_w(i,j)$ and $I'_w(i,j)$ are the original and the extracted watermark bits at pixel (i,j) respectively. Note that the maximum value of NC is 1. Also, the higher the NC value the more accurate the extracted watermark will be.

For the experimental results, we first compared the accuracy of the extracted watermark. The results in terms of average NC at various qualities of watermarked image from the proposed method and the previous method [5] with $\sigma = 3$ were measured, and illustrated in Fig. 3. It is obvious that the performance of the proposed method was superior to the previous method. Fig. 4 shows the some results of the watermarked images and the corresponding extracted watermarks from the two watermarking methods at PSNR of 35±0.01 dB.

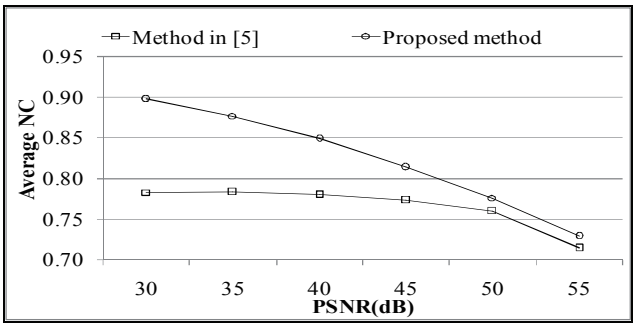

Fig. 3. Performance comparison between the two watermarking methods at various PSNR values

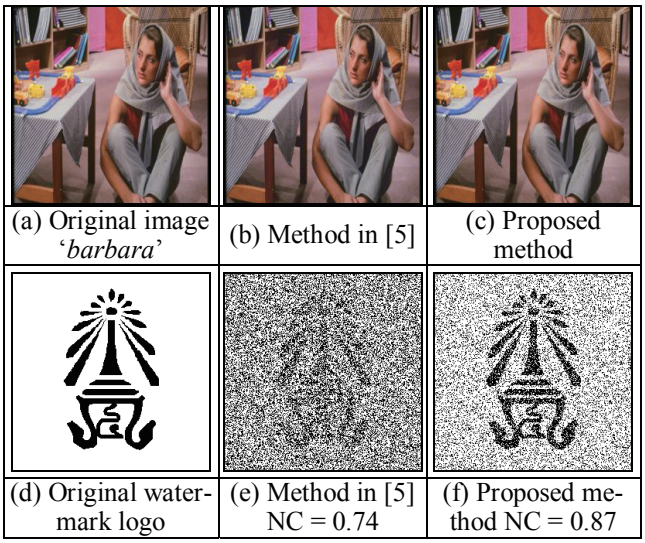

Fig. 4. (a) Original image '*barbara*' and (b), (c) its different watermarked versions (d) original watermark logo and (e), (f) its corresponding extracted watermark

Note that the methods in [5] embed the watermark in accordance with the standard deviation of the embedding components, some watermark bits would be omitted and discarded. When extracting a watermark bit from the non-embedded component, the method would assign a random bit/value, so that the extracted watermark has the same size as its original and could be used in the calculation of NC. Next, we evaluated the robustness of the embedded watermark by applying different types of attack at various strengths to watermarked image, obtained from the two methods at PSNR of 35±0.1 dB. Seven types of attacks was considered and used in the experiments, i.e. JPEG compression standard at various image qualities, the zero mean additive Gaussian distributed noise at various variances, brightness enhancement at various percentages, contrast adjustment at various scaling factors, image blurring at various pixel values, image rescaling at various ratios, and salt & pepper noise adding at various densities, as indicated in Fig. 5-11, respectively. Also, in the experiments, we did not perform the watermark resynchronization by using an image registration technique as described in [5]. We extracted the watermark directly from the attacked watermarked image. This is because we wished to evaluate and compare the true performance of the two watermarking methods where watermark resynchronization loss existed.

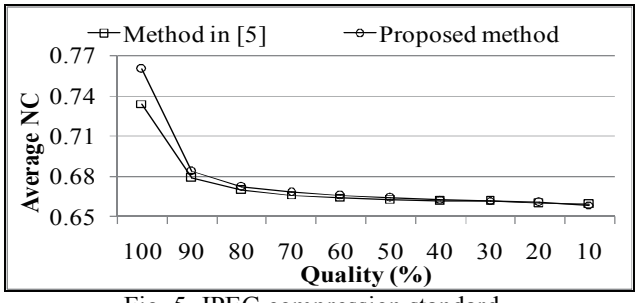

Fig. 5. JPEG compression standard

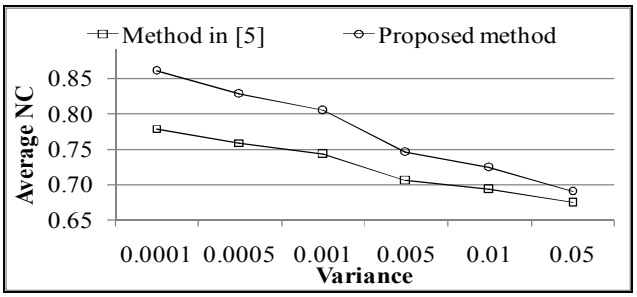

Fig. 6. Zero mean Gaussian noise adding

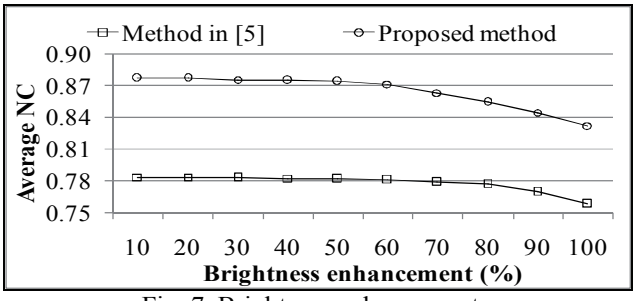

Fig. 7. Brightness enhancement

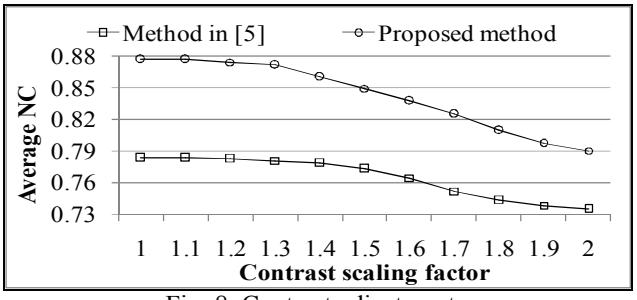

Fig. 8. Contrast adjustment

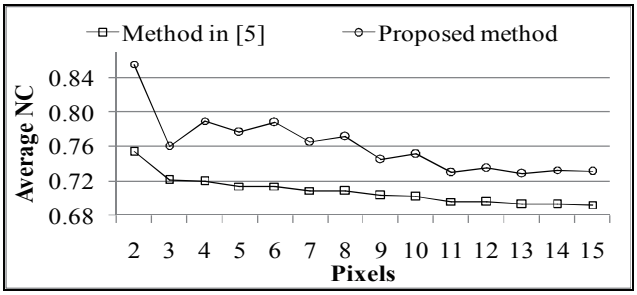

Fig. 9. Image blurring

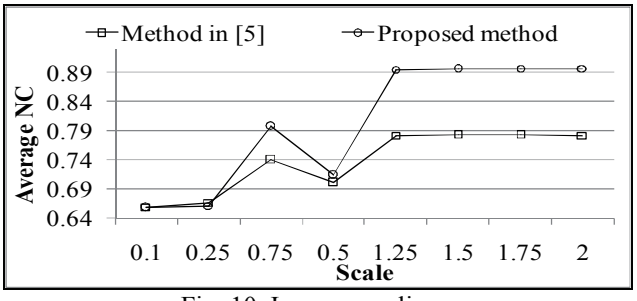

Fig. 10. Image rescaling

According the results obtained, it can be summarized that the robustness of the proposed method was stronger than the previous one, judged from a higher average NC value at the equivalent PSNRs. Fig. 12 shows some results of the extracted watermark after being attacked by image blurring at 4 pixels, contrast adjustment at scaling factor of 1.9, brightness enhancement at 100 percentages, and Gaussian noise at variance of 0.001, respectively.

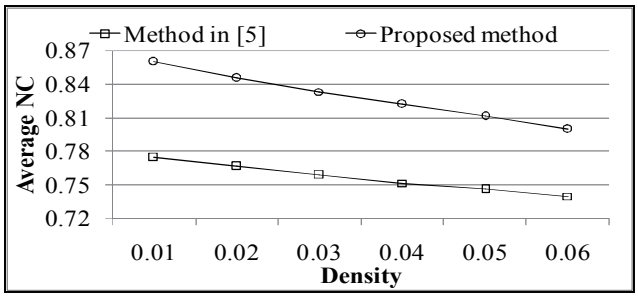

Fig. 11. Salt & pepper noise adding

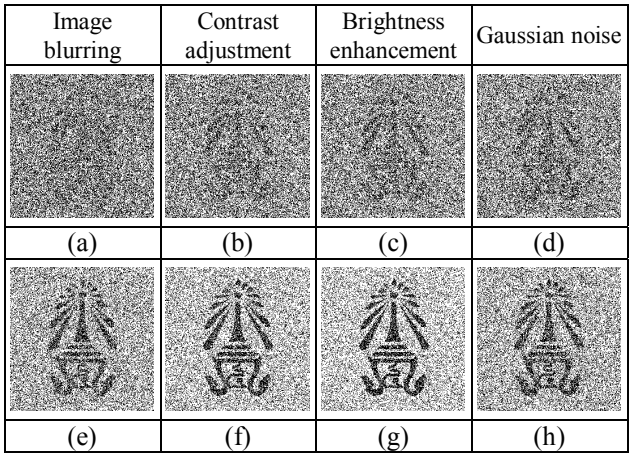

Fig. 12. (a)-(d) I'_w from the method in [5] and (e)-(h) I'_w from the proposed method

4. Conclusion

We presented in this paper a blind image watermarking based on regularized filter for color images. We experimentally demonstrated that the proposed method achieved higher accuracy of the extracted watermark at equivalent quality of the watermarked images, compared to the previous watermarking method in [5].

References

[1] Y. Wang, et al.: "A Wavelet Based Watermarking Algorithm for Ownership Verification of Digital Images," *IEEE trans. Image Processing*, vol. 11, no. 2, pp. 77-88, 2002.

[2] M. A. Suhail, et al.: "Digital Watermarking Based DCT and JPEG Model," *IEEE trans. on Instrumentation and Measurement*, vol. 52, no. 5, pp. 1640-1647, 2003.

[3] M. Kutter, et al.: "Digital signature of colour images using amplitude modulation," *Journal of Electronic Imaging*, vol. 7, no. 2, pp. 326-332, 1998.

[4] T. Amornraksa, et al.: "Enhanced images watermarking based on amplitude modulation," *Image and Vision Computing*, vol. 24, no. 2, pp. 111-119, 2006.

[5] C.-H. Lai, et al.: "Robust Image Watermarking Against Local Geometric Attacks Using Multiscale Block Matching Method," *Journal of Visual Communication and Image*, vol. 20, no. 6, pp. 377-388, 2009.

[6] J. A. Hussein, "Spatial domain watermarking scheme for colored images based on log-average luminance," *Journal of Computing*, vol.2, no.1, pp. 100-103, 2010.