Adaptive semi-fragile watermarking based on complete quantization and image contents

Wu Guo, Wei Dawei , Tang Guangming , Yang Jinfeng
Zhengzhou Information Science and Technology Institute, Zhengzhou, China
kimi_liujing@163.com , tgm1983@sina.com , muyue1983@163.com , yjfeng1983@126.com

***Abstract*— Complete quantization principle is proposed in this paper, based on which, an adaptive semi-fragile watermarking algorithm based on image contents is proposed. The algorithm uses the low frequency sub-band to construct watermarking related to the content of image, and then uses just noticeable difference of wavelet domain to embed watermarking by means of complete quantitative principle. Experiments show this algorithm can classify malicious and non-malicious manipulations accurately.**

Keywords- semi-fragile watermarking; quantization; image contents; wavelet coefficients

I. INTRODUCTION

With the development of network and multimedia technology, it's convenient to copy and tamper the digital works. Contents protection becomes one of the most urgent issues. Fragile and semi-fragile watermarking is effective technique to deal with content protection. Fragile watermarking can detect all the changes of digital works. However, since the slight signal processing operations caused by transmission and storage, like JPEG compression, noise and filter, are considered acceptable or necessary, semi-fragile is more suitable for practical applications.

Compared with semi-fragile algorithm based on fixed quantization step[1][2], the algorithm with dynamic quantization step can balance the robustness and imperceptibility better. Che[3] used human visual characteristics to determine the steps dynamically and proposed quantized central limit theorem to embed watermark. However, when the value of cover data is smaller than that of step, the parity of watermarking bit will be changed even if the cover data isn't modified. In addition, the watermarking of most semi-fragile algorithm is irrelevant to the cover image[3][4], so a reference watermarking is required when authenticating, which will increase the authentication risk.

In this paper, a novel semi-fragile watermarking method is proposed. Complete quantization principle presented in the paper makes sure the correction of watermarking extraction if the change of cover data belongs to a certain range. And then the watermarking is generated according to image's contents, therefore, the reference watermarking can be extracted from the unknown cover and " one image, one watermark" can be achieved, which reduces the risk of image block replacement attack compared with "many image, one watermark". Considering that texture and edge characteristics locate at high-frequency sub-band after wavelet decomposition[5], the embedding domain of the new method is HH_1.

II. COMPLETE QUANTIZATION PRINCIPLE

Let x be cover data and Δ be quantization step, then $x/\Delta = \mathrm{Q} + r$, where r is the remainder and has the some sign with x . Let w be the watermarking bit, x' be the data after embedding, then the complete quantization principle can be described as:

Complete quantization principle 1: when $\mathrm{mod}(\mathrm{Q},2) = w$:

If $x \geq 0$, then $x' = x + \Delta/2 - r$;

If $x < 0$, then $x' = x - \Delta/2 - r$.

Complete quantization principle 2: when $w = 0$ and $\mathrm{mod}(\mathrm{Q},2) = 1$:

If $x < 0$ and $r \geq \Delta/2$, $x' = x + 3\Delta/2 - r$;

If $x \geq 0$ and $r < \Delta/2$, $x' = x - \Delta/2 - r$;

If $x < 0$ and $|r| \geq \Delta/2$, $x' = x - 3\Delta/2 - r$;

If $x < 0$ and $|r| < \Delta/2$, $x' = x + \Delta/2 - r$.

Complete quantization principle 3: when $w = 1$ and $\mathrm{mod}(\mathrm{Q},2) = 0$:

If $x \geq 0$ and $r \geq \Delta/2$, $x' = x + 3\Delta/2 - r$;

If $x \geq 0$ and $r < \Delta/2$, $x' = \begin{cases} x + 3\Delta/2 - r & x < \Delta \\ x - \Delta/2 - r & x > \Delta \end{cases}$;

If $x < 0$ and $|r| \geq \Delta/2$, $x' = x - 3\Delta/2 - r$;

If $x < 0$ and $|r| < \Delta/2$, $x' = \begin{cases} x - 3\Delta/2 - r & x < \Delta \\ x + \Delta/2 - r & x > \Delta \end{cases}$.

Theorem 1: After adjusting the cover data using step Δ according to complete quantization principle, if the change of adjusted data belongs to $(-\Delta/2, \Delta/2)$ after processing or attack, the watermarking can be detected correctly.

Proof: According to complete quantization principle 1, when $\mathrm{mod}(\mathrm{Q},2) = w$, the adjusting scheme of cover data is shown as Fig. 1 (a). Namely, x is modulated to the center of its quantization interval. Therefore, when the change of x' belongs to $(-\Delta/2, \Delta/2)$, the parity of quantized data can't be changed and the watermarking can be detected correctly.

Fig.1 (b) and Fig.1 (c) show the adjusting schem of complete quantization principle 1 and 2 respectively. The theorem can be proved in the same way.

Published by Atlantis Press, Paris, France.

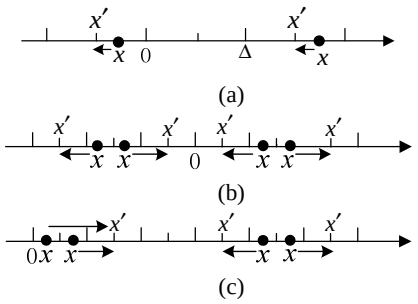


Figure 1. The adjusting scheme of cover data.

III. WATERMARKING GENERATION BASED ON CONTENT

The watermarking is generated in the low-frequency sub-band LL_3 of wavelet decomposition with scale three in order to make it fragile to malicious attack. The generating process is shown below:

Step1: For one image I sized $M \times N$, the low-frequency sub-band LL_3 is obtained after Haar wavelet decomposition with scale three.

Step2: The binary watermarking matrix is generated by :

$$w(i,j)=\begin{cases}1 & LL_3(i,j)\ge\beta\\ 0 & LL_3(i,j)<\beta\end{cases}$$

Where $\beta = Max(|LL_3|)/2$, $LL_3(i,j)$ is the wavelet coefficient of LL_3.

IV. ADAPTIVE DETERMINATION OF QUANTIZATION STEP

According to human visual characteristics, the imperceptibility of watermarking is influenced by texture complication, background luminance and frequency of embedding domain[6].

- Texture masking factor is:

$$\Xi(l,i,j)=\sum_{k=0}^{3-l}\frac{1}{16^k}\sum_{\theta=0}^{2}\sum_{x=0}^{1}\sum_{y=0}^{1}\left[f_{k+l,\theta}(y+\frac{i}{2^k},x+\frac{j}{2^k})\right]^2\cdot Var\left\{f_{3,3}(1+y+\frac{i}{2^{3-l}},1+x+\frac{j}{2^{3-l}})\right\}_{\substack{x=0,1\\y=0,1}}$$

where l is the scale, (i,j) denotes the coordinate of coefficient, $\theta=0,1,2,3$, is the sub-band of LL、HL、LH and HH. $f_{l,\theta}(i,j)$ is the wavelet coefficient at (i,j) in θ subband with scale l.

- Luminance masking factor is $\Lambda(l,i,j)$, then:

$$\Lambda(l,i,j)=1+L'(l,i,j)$$

Where $L'(l,i,j)=\begin{cases}1-L(l,i,j), & L(l,i,j)<0.5\\ L(l,i,j), & otherwise\end{cases}$

$$L(l,i,j)=\frac{1}{256}I_3^3(1+\left\lfloor\frac{i}{2^{3-l}}\right\rfloor,1+\left\lfloor\frac{j}{2^{3-l}}\right\rfloor)\cdot$$

- Noisy masking factor is $\Theta(l,\theta)$, then it can be estimated by:

$$\Theta(l,\theta)=\begin{cases}\sqrt{2} & if\quad \theta=1\\ 1 & otherwise\end{cases}\cdot\begin{cases}1.00, & l=0\\ 0.32, & l=1\\ 0.16, & l=2\\ 0.10 & l=3\end{cases}$$

Therefore, the just noticeable difference (JND) of $f_{l,\theta}(i,j)$ is:

$$JND_{l,\theta}(i,j)=\Xi(l,i,j)\,\Lambda(l,i,j)\Theta(l,\theta)/2$$

A block quantization scheme is adopted in this paper for improving the robustness to conventional processing such as JPEG compression. The HH_1 sub-band is divided into 2×2 blocks and one bit is embedded into one block. Therefore, the JND of wavelet coefficient is required to turn to that of image block and then the quantization step can be determined. What's more, the step can't be influenced by embedding for blind detection. Since there is relevance among neighbor sub-bands, the step of each block can be calculated by the block JND of the same location in HL_1 and LH_1:

$$\Delta_1^{\theta=3}(k)=\ln\left(\left(\frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2}\left|JND_{1,\theta=1}^{k}(i,j)\right|+\frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2}\left|JND_{1,\theta=2}^{k}(i,j)\right|\right)\Big/2\right)$$

Where, $\Delta_1^{\theta=3}(k)$ is the quantization step of the k th block in HH_1, $JND_{1,\theta=1}^{k}(i,j)$ and $JND_{1,\theta=2}^{k}(i,j)$ are the JND of coefficient at (i,j) in k th block of HL_1 and LH_1 respectively.

V. WATERMARKING ALGORITHM

A. Watermarking embedding algorithm

Step1: Image I is decomposed by Haar wavelet. The watermarking is generated using the method in section 3 and HH_1 is divided into block sized 2×2.

Step2: Select a key and generate the embedding path by random generation.

Step3: Calculate the coefficient mean of k th block: $D(k)=\sum_{i=1}^{2}\sum_{j=1}^{2}f^k(i,j)\big/4$.

Step4: Calculate $D(k)/\Delta(k)=Q(k)+r(k)$, where $\Delta(k)$ is the quantization step obtained by the method in section 4.

Step5: Modify the coefficient of k th block $f^k(i,j)$ to embed watermarking, and the new coefficient $\tilde{f}^k(i,j)$ is got.

Step6: Substitute $f^k(i,j)$ with $\tilde{f}^k(i,j)$, and combine it with the coefficient without watermarking. The image with watermarking can be obtained after inverse wavelet decomposition.

B. Watermarking extracting algorithm

Step1: The unknown image $\tilde{I}$ is decomposed by Haar wavelet and sub-band HH_1 is divided into blocks sized 2×2.

Step2: Extract the image blocks with watermarking according to the key.

Step3: Calculate the coefficient mean of above block $D(k),1\le k\le MN/64$, and $D(k)$ is divided by the quantization step of that block and $Q(k)$ is obtained.

Step4: Calculate $\hat{w}(k)=\mathrm{mod}(Q(k),2)$, then the watermarking bit is extracted.

C. Watermarking authentication

According to the theory of Kundur[7], the processing result of image is the Gaussian distribution with mean 0 and variance σ^2 of wavelet coefficient. σ^2 relates to Conventional processing is smaller than that of malicious

attack. Therefore, the type of image processing can be defined by the block temper rate between reference and extracted watermarking. The equation is :

$$r(i', j') = \frac{\sum_{i=(i'-1)\times m+1}^{i'\times m} \sum_{j=(j'-1)\times n+1}^{j'\times n} d(i, j)}{m \times n}$$

Where $1 \le i \le (M/8), 1 \le j \le (N/8)$, $d(i, j) \in \hat{W} \oplus W'$, $m \times n$ is the size of block.

Obviously, for conventional processing ,the error point is disperse, so $r(i', j')$ is small and vice visa. The type of image processing can be decided by setting a detection threshold T :

- If $\max(r) = 0$, the image isn't processed.
- If $\max(r) \le T$, the image is processed by conventional processing without tamper.
- If $\max(r) > T$, the image is tampered by malicious attack.

The detection threshold T is determined by the quality of image. When a higher quality of image is required, the value of T is smaller.

VI. EXPERIMENT RESULTS AND ANALYSIS

800 gray-scale images[8] with fixed size 512×512 are employed to evaluate our algorithm. The threshold $T = 0.45$ and $m = n = 8$.

A. Experiment of conventional image processing

In order to test the robustness of our method, the image with watermarking is processed by JPEG compression, noise adding, gray enhancement and filtering, then the watermarking is extracted and detected. Take image "Lena" for an example, Table 1 shows the detecting results of conventional processing.

TABLE I. DETECTING RESULTS OF CONVENTIONAL PROCESSING

Processing type	Parameter	$\max(r)$	Result
JPEG compression	100	0.000	untemperd
	90	0.094	untemperd
	70	0.125	untemperd
	40	0.343	untemperd
Noise adding	Peppr&salt noise 1%	0.141	untemperd
	Peppr&salt noise 2%	0.250	untemperd
	Peppr&salt noise 3%	0.375	untemperd
	Gaussian noise 20dB	0.047	untemperd
	Gaussian noise 35dB	0.313	untemperd
	Poisson	0.188	untemperd
Gray enhancement	[0.15，0.9]，[0，1]	0.391	untemperd
Filtering	Log filtering 3×3	0.421	untemperd
	Median filtering 2×2	0.375	untemperd
	Median filtering 3×3	0.406	untemperd
	Median filtering 2×2	0.250	untemperd
	Median filtering 3×3	0.344	untemperd

After all the processing shown in above table, the values of max(r) of all the extracted watermarking are smaller than the detecting threshold, indicating the content of image isn't tampered. Therefore, the algorithm is robust to conventional image processing.

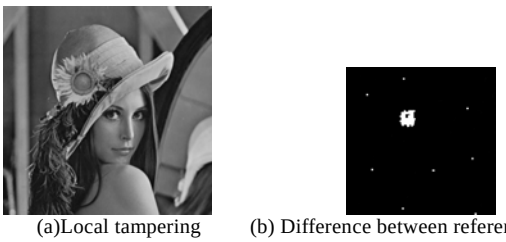
B. Experiment of malicious tamper

In order to test the sensitivity and the location capability of our method, the images are processed by cropping and local tampering. Take image "lena" as an example, the experiment results are shown below.

Table 2 shows detection results of malicious tamper. It can be seen that the values of max(r) of all the extracted watermarking are larger than the threshold T , indicating the content of image is tampered.

(a)Cropping (b)Difference between reference and extracted watermarking

Figure 2. Result of cropping.



(a)Local tampering (b) Difference between reference and extracted watermarking

Figure 3. Result of local tampering.

TABLE II. RESULTS OF MALICIOUS TAMPER

Attack type	$\max(r)$	result
Cropping	0.922	tampered
Local tampering	0.969	tampered

From Fig.2 and Fig.3, the location of tampered content can be oriented well according to the difference between

reference and extracted watermarking. Therefore, the algorithm is sensitive to malicious tamper and can orient the location of tampered content well.

C. Experiment of comparing with other methods

We compare our method with the methods in [3] and [9], which perform better than most of semi-fragile watermarking algorithm. Table 3 shows the results of false alarm rate and missing alarm rate.

TABLE III. RESULTS OF FALSE ALARM RATE AND MISSING ALARM RATE

Method	Missing alarm rate of 1/16 cropping (%)	False alarm rate of conventional image processing (%)					
		Without processing	JPEG compression (70%)	Gaussian noise (35dB)	Pepper &salt noise (1%)	Gray enhancement	Median filtering (2×2)
Method in [9]	0.5	3.6	5.2	0.3	8.7	6.5	7.1
Method in [3]	0.0	0.6	2.6	0.7	3.4	4.2	3.2
Our Method	0.0	0.5	2.4	1.0	2.9	3.8	3.0

Form the results of table 3, the performance of our method is better than that of [9] and close to that of [3]. However, the watermarking of [3] is unconcerned with the image, so the original watermarking is required when authenticating.

VII. CONCLUSIONS

The watermarking related with image's content is generated in low frequency sub-band of wavelet decomposition, so the watermarking is fragile to malicious tamper and robust to conventional processing and the original watermarking isn't required when authenticating. The imperceptibility of watermarking embedding is achieved by adaptively adjusting quantization step. Complete quantization principle is proposed to modulate the mean of wavelet coefficient block. The watermarking bit can be extracted without additional information and recovered correctly when the change of image belongs to a certain range.

REFERENCES

[1] Ekici Ö, Sankur B, Akcay M. Comparative assessment of semifragile watermarking methods[J]. Journal of Electronic Imaging, 2004,13(1): 209-216.

[2] Zhang Jing, Zhang Chuntian. Digital watermarking techniques for image authentication[J]. Journal of Image and Graphics 2003,8(4):367-373.

[3] Che Shengbing, Huang Da, Li Guang. Semi-fragile image watermarking algorithm based on visual features[J]. Journal on Communications, 2007, 28(10): 134-140.

[4] Tang qian. Study on semi-fragile watermarking algorithm based on edge detection[J].Computer engineering, 2011, 37(11):178-180.

[5] Lewis A S, Knowles G. Image compression using the 2-D wavelet transformation[J]. IEEE Transactions on Image Processing, 1992, 1(4): 244-250.

[6] Wang Xiangyang, Yang Hongying, Chen Like, Zhao Hong. A new semi-fragile image watermarking based on visual masking[J]. Journal of Image and Graphics, 2005, 10(12): 1548-1553.

[7] Kundur D, Hatzinakos D. Towards a telltale watermarking technique for tamper proofing[A]. Proceedings of the IEEE International Conference on Image Processing[C], Chicago, USA, 1998, 2: 409-413.

[8] Li J. Photo graphy image database[EB/OL]. http://www.stat.psu.edu/jiali/index.download.htm.

[9] Gopalakrishnan, T.; Ramakrishnan, S.; Balasamy, K.; Murugavel, A.S.M.. Semi fragile watermarking using Gaussian mixture model for malicious image attacks [C]. 2011 World Congress on Information and Communication Technologies,2011: 120 – 125.

Published by Atlantis Press, Paris, France.