Adaptive attacking algorithm against DCT-based watermarking

Tao Zhang[1] Daoshun Wang[1] Shundong Li[2] Xunxue Cui[3] Yiqi Dai[1]

[1]Tsinghua National Laboratory for Information Science and Technology (TNlist), Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, P.R.China
[2]School of Computer Science, Shaanxi Normal University, Xi'an 710062, P.R.China
[3]New Star Research Institute of Applied Technology, Hefei, 230031, P.R.China

Abstract

In this paper we present a new watermarking attacking algorithm based on the periodic transformation of matrix. By analyzing of the principles of adaptive watermarking embedding in DCT domain, we chose some blocks of the stegoimage to embed the watermarking based on the characteristics of human visual system (HVS), and perform the periodic transformation to the chosen blocks. If we perform periodic transform to the same blocks certain more times, the original stegoimage can be recovered. So we can prove the validity of our attacking method to the owner of watermarking system.

Keywords: watermarking attack, periodicity, matrix transformation, DCT, HVS.

1. Introduction

The direct target of watermarking attack is to destroy the watermarking or prevent it from recovering. Watermarking attack can test the performances of the watermarking system, as well as push forwards the development of watermarking technology and design some new better methods. Many scholars have done much constructive research work. For instance, Petitcolas [1, 2] *et al*. consider that the watermarking attacking methods can be divided into three categories, which are basic attack(including jitter attack, synchronization attack and subtle distortion attack), stirmark attack and interpretation attack.

Many kinds of benchmark software have been developed to test the watermarking attacking methods based on the studies of watermarking attack, which include Stirmark, Heckmark, Ertimark [4]. Petitcolas [5] *et al.* present an architecture of a public automated evaluation service developed for still images, sound and video which is a series of tests applied to different types of watermarking schemes.

We can divide all attacking methods into two categories according to whether the quality of the carrier is obviously debased. The first category of attacking methods obviously debates the quality of the carrier. Geometrical attack and image processing attack [1, 2, 6, 7] are the representative methods in the first category. This category of attacking methods and corresponding benchmark have become the models of authority, so it is hard to do any breakthrough in this area. And the second category of attacking methods does not appreciably debate the quality of the carrier. Mosaic attack [1, 2, 3], watermark template attack [6] and adaptive attack [7] are the representative methods in the second category. There are some problems to be solved in the second category attacking methods. For instance, the

carrier contains intact watermark after the mosaic attack which can be detected. In other words, we can not prove the success of mosaic attack if the carrier is integrated. What's more, if the watermark needs to be recovered after mosaic attack, watermark template attack and adaptive attack, some complex adverse transforms are necessary which need much more additive information and can not be completed automatically.

This thesis analyzes the principles of adaptive watermarking embedding, and adopts the periodic matrix transformation to attack the areas in the stegoimage which are chosen according to the characteristics of HVS. Also we do some experiments using the new method in DCT domain, which show that the new method has some perfect effects. For example, it could prevent the detecting and recovering of the watermarking and ensure the quality of the stegoimage, as well as recovering the original stegoimage by attacking the same image in the same way some more times.

2. The analyses of watermarking embedding area in DCT domain

2.1. Discrete Cosine Transform (DCT)

DCT is a typical digital image transformation. As we know, a digital image can be seen as the sample values of dualistic function in discrete grids, so it can be denoted as matrix. Besides, the contents of an image are usually self-correlative. That is, the contents of local image sometimes change little. If we take an image as the functional value in same distance grid of a continual dualistic function $F(x, y)$, to the given $\varepsilon > 0$, $\Delta x > 0$, $\Delta y > 0$, we have $|F(x, y) - F(x', y')| < \varepsilon$, $x \in [x' - \Delta x, x' + \Delta x]$, $y \in [y' - \Delta y, y' + \Delta y]$.

By using DCT, we can make use of self-correlation of digital image to reduce information, so the digital can be compressed. The popular digital image and video compression standards such as JPEG and MPEG use DCT as their cores.

The two dimension DCT is defined as follow,

$$G_c(u,v) = a(u)\,a(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}(g(x,y)\times \cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (2.1.1)$$

And its inverse transformmation is as follow,

$$g(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}(a(u)\,a(v)\,G_c(u,v)\times \cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (2.1.2)$$

Here, $a(0) = \sqrt{\frac{1}{N}}$ and $a(m) = \sqrt{\frac{2}{N}}$, $1 \le m \le N$.

2.2. Human Visual System (HVS)

It is popular to adaptively embed the watermarking in the DCT domain based on the characteristics of HVS, so it is necessary to analyze corresponding policies of adaptive embedding methods.

HVS has three distinct characteristics. First, it has different sensitivity against different grey scale. The sensitivity is strongest when the grey scale is medium and nonlinearly declines when grey scale become higher or lower. Second, it is sensitive against smooth areas and not sensitive against texture areas of the image. Third, it is very sensitive against borderlines of the image. Generally, if the intensity of signals is under sensitivity threshold of the HVS, HVS can not feel the existence of the signals [8].

We divide the stegoimage into many sub-blocks by 8×8 pixels and calculate the entropy and variance of each sub-block. The sub-block should be smooth area if the entropy is high and should be

texture of borderline area if the entropy is low. The variance is smaller in texture area and bigger in borderline area. We can divide all sub-blocks of the image into 4 categories according to the characteristics of HVS and adopt corresponding policies when adaptively embed watermarking in DCT domain [9].

1) The first category. The sub-blocks in this category have low luminosity and simple texture, in which HVS is considerable sensitive to the changes of pixels. The intensity of embedded watermarking should be the lowest.

2) The second category. The sub-blocks in this category have high luminosity and complex texture; however, they are in borderline areas. The intensity of embedded watermarking should be considerable low.

3) The fourth category. The sub-blocks in this category have high luminosity and complex texture; what's more, they are not in borderline areas. HVS is the most insensitive to the changes of pixels in these sub-blocks. The intensity of embedded watermarking should be the highest.

4) The third category. The other sub-blocks except the three categories above can be called the third category.

The thresholds to classify the sub-blocks are decided by experiments.

3. The Periodic Matrix Transformation

Any digital image has a corresponding numerical value matrix, and a pixel in the image is corresponding to an element in the matrix. The various image processing is equal to various matrix transformations. If the elements values in the matrix are changed, the corresponding image is also changed to another one.

Some of image scrambling methods have interesting periodic characteristic, such as Arnold, Fibonacci-Q and A-F methods. A common periodic matrix transformation can be abstracted from these methods [10].

In the sequel, for convenience, let $\mathbf{X}_n = (x_1, ..., x_n)^T$, $\mathbf{X}_n' = (x_1', ..., x_n')^T$, $x_1, ..., x_n \in \{0, 1, ..., N-1\}$.

Definition 1: For an arbitrarily given positive integer N and a digital image $\mathbf{P}$, the following transformation

$$\mathbf{X}_n' = \mathbf{A}\mathbf{X}_n \pmod N \quad , \quad (\ \mathbf{A} = (a_{ij})_{n\times n} \ , \ a_{ij} \in \mathrm{z}\ ) \tag{3.1}$$

has a period m_N with respect to the image $\mathbf{P}$ and m_N is the minimal times that make the image $\mathbf{P}$ return to its original status. For an arbitrary matrix $\mathbf{A} = (a_{ij})_{m\times n}$, we have $\mathbf{A} \pmod N = (a_{ij} \pmod N)_{m\times n}$.

Proposition 1: For a given fixed positive integer N , if $\mathbf{X}_n' = \mathbf{A}\mathbf{X}_n \pmod N$, $\mathbf{A}^m \mathbf{X}_n \pmod N$ is obtained after m times transformations for $\mathbf{X}_n$.

Using proposition 1, the following results can be obtained.

Proposition 2: If Transformation (3.1) has the period m_N , then m_N is the smallest positive integer which makes $\mathbf{A}^{m_N} (\mathrm{mod}\, N) = \mathbf{E}_n$, where $\mathbf{E}_n$ is the n -order unitary matrix.

Theorem 1 [10]: The sufficient and necessary condition that transformation (3.1) has the periodicity is that $|\mathbf{A}|$ and N are prime to each other, where $|\mathbf{A}|$ is the determinant of the matrix $\mathbf{A}$.

In the following, we give a periodic transformation matrix that satisfies theorem 1.

$\mathbf{X}_n' = \mathbf{A}_n \mathbf{X} (\mathrm{mod}\, N) =$

$$\begin{bmatrix} 2 & 3 & 2 & 2 & 2 & 2 & \dots & 2 & 2 & 2 \\ 3 & 4 & 3 & 4 & 4 & 4 & \dots & 4 & 4 & 4 \\ 4 & 5 & 3 & 5 & 6 & 6 & \dots & 6 & 6 & 6 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ n-1 & n & 3 & 5 & 7 & 9 & \cdots & 2n-7 & 2n-5 & 2n-4 \\ n & n+1 & 3 & 5 & 7 & 9 & \dots & 2n-7 & 2n-5 & 2n-3 \\ n+1 & n+2 & 3 & 5 & 7 & 9 & \dots & 2n-7 & 2n-5 & 2n-3 \end{bmatrix} \mathbf{X} (\mathrm{mod}\, N)$$

The periods of $\mathbf{A}_n$ are shown in Tab. 1 when 2 and 3 dimension transformations have different rank N .

Tab. 1: The periods of different dimensional transformation relevant to N

Dimension	Periods							
	N=2	3	4	5	6	7	8	9
2	2	3	4	20	2	16	8	6
3	7	3	7	10	21	24	14	9
	N=10	20	63	64	124	128	255	256
2	20	29	48	64	60	128	180	256
3	70	70	72	112	2317	224	9210	448

The following example shows how to calculate the period.

$\mathbf{A}_2 = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}$ and $N = 5$, $\mathrm{mod}([\mathbf{A}_2 \times \mathbf{A}_2 \times \cdots \times \mathbf{A}_2]^{20}, 5) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, from Proposition 2, the period of matrix $\mathbf{A}_2$ is 20.

4. Attacking algorithm

Commonly, according to the characteristics of HVS, the second, third and fourth categories blocks mentioned in section 2 are chosen to embed watermarking, so some categories blocks should also be chosen to attack. The processes in the frame hereinafter are the attacking algorithm details.

Despite any matrix meeting the conditions mentioned in section 3, the matrix with the small period and ±1 relatively prime should be better, which could make the calculation easier and the elements value positive in inverse transformation matrix. Fig. 1 and Fig. 2 show the process of detecting watermarking in attacked stegoimage and recovered stegoimage.

Input:: stegoimage
Output: attacked image
Attack algorithm:
(1) Divide the stegoimage into sub-blocks by 8×8 pixels, do DCT transform and on every sub-block and calculate which category the sub-block should belong to.
(2) k=1, m=the threshold times of transform.
(3) Choose some pixels in the second, third and fourth categories sub-blocks and construct a column vector $\mathbf{B}_i$, choose a periodic matrix $\mathbf{A}$, do the transform $\mathbf{B}_i' = \mathbf{A}\mathbf{B}_i (\mathrm{mod}\, N)$.
(4) Replace $\mathbf{B}_i$ by $\mathbf{B}_i'$, do IDCT transform.
(5) calculate PSNR，if PSNR>28 and k<m , k=k+1, goto（3）.

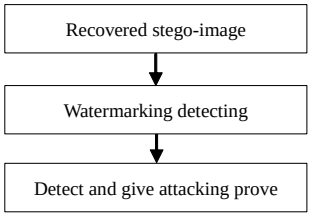

Fig. 1: Detect after attack

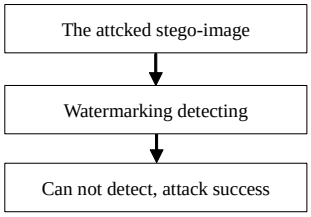

Fig. 2: Detect after recover

5. Experimental results

We adopt the international standard software CoxWM to embed watermarking into the image in DCT domain, and attack the stegoimage time after time with the same algorithm. Because the PSNR is infinity when the stegoimage is recovered, we can observe the periodic characteristic of the attacking algorithm through the value changes of the PSNR. At the same

time, we detect watermarking in the stegoimage after every time of attack.

We adopt periodic matrix transformation with dimension n ($n=2,3$) and modulus N (the value of N is commonly the biggest value of pixel in the input image) to do the experiment. In this paper, we use N to present the largest position of pixels to be transformed in $n \times n$ DCT sub-block. The periodic matrix can be the $n \times n$ up-left sub-block of the matrix mentioned in section 3, whose period is shown in Tab. 1. We can not detect watermarking in the stegoimage when attacking times $m < m_N$; and can detect the same watermarking as before attack when attack times m is the integral times of m_N , at the same time, we can observe that the PSNR is infinity which shows the stegoimage is completely recovered. The experimental results are shown in Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Tab. 2 and Tab. 3. Because the DCT sub-block is divided by 8×8 pixels, we set N=7.

We can compare the experiment results with Stirmark(version 3-1-79, and the parameters are default). The results of Stirmark attack are shown in Fig. 8 and Fig. 9, where we can observe the stegoimage is obviously blurred and distorted after Stirmark attack, which the PSNR is 18.7302. The effect of our attack algorithm is better than Stirmark attack algorithm.

Fig.3: stegoimage

Fig. 4: stegoimage after attack when n=2 and m=5

Fig. 5: stegoimage after attack when n=2 and m=20

Fig. 6: stegoimage after attack when n=3 and m=30

Fig. 7: stegoimage after attack when n=3 and m=70

Tab. 2: The changes of PSNR after different times of attack

Dimension	SNR					
	Aattack times =1	2	3	4	5	8
2	43.37	43.17	43.22	43.02	43.34	43.28
3	42.36	42.19	42.27	42.11	42.29	_
	Aattack times =15	20	30	40	50	70
2	43.23	∞	42.92	∞	42.92	42.92
3	_	_	42.45	42.18	42.25	∞

Tab. 2: PSNR, correlation factor and conform factor

comparing objects	PSNR	CF	Similitude
Cox wateramrking image versus original image	31.99	0.99	29.82
image attacked once versus Cox watermarking image	43.37	6.38E-5	0.23
recovered image versus Cox watermarking image	INF	1	1

Fig. 8: Original watermarking image

Fig. 9: Watermarking image after Stirmark attack

6. Conclusion

In this paper we bring forward a new watermarking attacking method based on HVS and periodic matrix transformation. This periodic method has good attacking effect and can completely recover the original stegoimage. If the owner of stegoimage denies his or her ownership, this attack method can be used to recover the original image and detect the watermarking to prove that the owner's denying. So this method can be used in special areas such as copyright protection or military affairs.

7. Acknowledgements

This research was supported in part by the National Natural Science Foundation of the People's Republic of China under Grant No. 90304014, 60673065, 60873249, and 60773129, 863 Project of China under Grand 2008AA01Z419 and the Project funded by Basic Research Foundation of School of Information Science and Technology of Tsinghua.

8. References

[1] Petitcolas, F., A., P., Anderson, R., J., and Kuhn, M., G., "Attacks on copyright marking systems", *Portland Oregon: Springer-Verlag*, 1525, pp.219-239, 1998.

[2] Petitcolas, F., A., P., Anderson, R., J., and Kuhn, M., G., "Information hiding - a survey", *Proc. IEEE 87*, pp.1062-1078, 1999.

[3] Petitcolas, F., A., P., "Stirmark benchmark 4.0", http://www.petitcolas.net/fabien/ watermarking/stirmark/ , 2007.

[4] Daoshun, W., Jinghong, L., Lei, Z., and et al., "Evaluation based on invisible watermarking system", *Proc. SPIE 5241, The International Society for Optical Engineering*, pp.202-210 , 2003.

[5] Petitcolas, F., A., P., Steinebach, Martin, and et al. “A public automated web-based evaluation service for watermarking schemes: StirMark Benchmark”, *SPIE 4314*, pp.575 – 584, 2001.
[6] Herrigel, A., Voloshynovskiy, S. and Rytsar, Y., “The watermark template attack”, *Proc. SPIE 4314, The International Society for Optical Engineering*, pp.394-405, 2001.
[7] Kutter, M. and Petitcolas, F., A., P., “Fair evaluation methods for image watermarking systems”, *Electronic Imaging*, papers 9, pp.445-455, 2000.
[8] Andrew, B., W., “DCT quantization matrices visually optimized for individual images”, *SPIE 1913*, pp.202-216, 1993.
[9] Huang, J., W. and Shi Y.,Q., “Adaptive image watermarking scheme based on visual masking”, *Electronics Letters*, papers 34, pp.748-750, 1998.
[10] Dongxu, Q., Daoshun, W. and Dilian, Y., “Matrix Transformation of digital image and its periodicity”, *Progress of Natural Science*, papers 6, pp.1-8, 2001.