

On Collusion Attack for Digital Fingerprinting

Xin-Wei Li, Bao-Long Guo and Xian-Xiang Wu

Institute of Intelligent Control & Image Engineering
Xidian University
2 South Taibai Road, Xi'an, 710071, China
lixinwei_xd@gmail.com; blguo@xidian.edu.cn; wuxianxiang@163.com

Lei-Da Li

School of Information and Electrical Engineering
China University of Mining and Technology
Xuzhou, 221116, China
lileida@cumt.edu.cn

Received April 2011; revised November 2011

ABSTRACT. *Collusion attack is a challenge issue for digital fingerprinting, and it is very beneficial to study for fingerprinting design. Two models for linear and non-linear collusion attack are constructed from the angles of owner and colluder in the paper. Through theory and simulation analysis in fingerprints embedding domain and non-embedding domain some conclusions are obtained: (1) The attack effects in embedding domain and non-embedding domain are same for average collusion attack; (2) The attack effects in embedding domain is stronger than that in non-embedding domain for maximum attack, minimum attack and randneg attack; (3) The attack effects in embedding domain and non-embedding domain are similar for median attack, min-max attack and modneg attack; (4) The two models have the similar attacks effects in embedding domain. We have also studied the relationship between the quality of colluded image and collusion attack. Large experimental results prove the correctness of the received conclusions, which is very helpful for fingerprinting design and system analysis.*

Keywords: Digital Fingerprinting; Linear Collusion; Nonlinear Collusion; Embedding Domain; Non-Embedding Domain

1. Introduction. With the development of computer and internet technology, the copy and transmission of multimedia works become more and more convenient. However, the multimedia works needing to protect are illegally copied and transmitted. Digital watermarking can protect multimedia intellectual property effectively. The work holder embeds some special information into the works to mark them before distribution, which aim to maintain his right when disputing. Digital fingerprinting is a special kind of watermarking [1-4]. It embeds user Identity(ID) into works before distributing them to user, thus the user can be traced when he sells his copy illegally. Digital fingerprinting is an effective way to protect copyright. More and more peoples devote themselves to the studying.

However, several illegal users may come together to generate a new copy using their authorized copies. The new copy can remove or weak the embedded information to make them evade punishment. The process is called collusion attack. Collusion attack is a challenge and difficult issue to solve for fingerprinting. The main concern for a fingerprinting

system is the performance of anti-collusion attack. So studying the characteristics of collusion attack is very valuable for fingerprinting design and analysis.

The prior research mainly concern two aspects. One research is fingerprint coding [5-9]. The other is studying fingerprinting scheme and the performance of fingerprinting system [10-15]. The two aspects are both aiming at strong robust to resist collusion attack. When designing fingerprinting, many literature [9-16] constructed collusion model in fingerprint embedding domain for convenience and premised that the host vector had been known. However, the illegal users dose not know the detail of embedding domain and the position of coefficients usually, and they may likely collude with all coefficients participating in. So it is necessary to study collusion attack in fingerprint embedding domain and non-embedding domain and compare their differences.

Stone studied some kinds of linear and nonlinear collusion attacks [17], and stated that nonlinear collusion is more effective than average attack when the fingerprints follow uniform distribution. He also indicated that Gaussian fingerprints are more robust than uniform fingerprints to resist collusion. Zane studied Gaussian fingerprinting system comprehensively, and focused on its ability to resist average attack [8]. Zhao ea el. constructed nonlinear collusion attack model and studied the ability of Gaussian fingerprinting system resisting nonlinear collusion comprehensively [18]. Varna et al. extended the application field of Gaussian fingerprinting system to compressed images [15]. The above methods all constructed collusion model with the suppose that the components of host vector are clear for describing simply. But the embedding domain and the components constructed host vector are unknown to colluders. Chen et al. stated that nonlinear collusion attack in embedding domain is more effective than that in non-embedding domain [19]. However, the attack effect of specific nonlinear collusion was not studied specifically. Because the assumptions from the angles of owner and colluders are different, the collusion model should be different. Thus the difference of the collusion attacks in different domain needs to be compared.

Partial-component model and all-component model are constructed in the paper. Gaussian fingerprints are embedded into Discrete Cosine Transform(DCT) domain which is assumed to be embedding domain. The fingerprints are embedded with the two models in DCT domain, spatial domain, wavelet domain, and their effects of collusion attack are compared. Some conclusions are obtained through model analysis and experiments, whose correctness is proved by further experimental results.

The organization of the paper is as follows. Section 2 constructs partial-component collusion model and all-component collusion model from different angles. Section 3 analyzes the collusion models in theory. Three conclusions are obtained under all kinds of collusion attacks. Section 4 gives the simulations aimed at the conclusions. Large experimental results including detection probabilities and the quality of colluded images are also shown. Conclusions are shown in the final section.

2. Collusion model. Collusion attack is classified into linear collusion and nonlinear collusion according to the performance of collusion model. A representative collusion for linear collusion is average attack. There are some typical nonlinear collusion named minimum attack, maximum attack, median attack, minmax attack, modified negative attack, randomized negative attack separately [11,12,18]. They all constructed the collusion model as follows.

Assume the host vector as \mathbf{S} with length N , and there are M fingerprints in the fingerprint family with length N denoted by $\{\mathbf{f}_k\}_{k=1}^M$. The fingerprinted copy distributed to the k th user is $\mathbf{Y}_k^{(j)} = \mathbf{S}_k^{(j)} + \alpha \mathbf{f}_k^{(j)}$. Where α is a constant, j represents the j th component in each vector. Suppose there are K colluders to conspire. The colluder subset denotes

by $S_c = \{1, 2, \dots, K\}$. The nonblind detection is employed to assurance high detection probability. Thus the partial-component collusion model is represented as follows.

$$\begin{aligned}
\text{Average attack: } \mathbf{V}_{aver}^{(j)} &= (\sum_{K \in S_c} \mathbf{f}_k^{(j)})/K \\
\text{Minimum attack: } \mathbf{V}_{min}^{(j)} &= \min(\{\mathbf{f}_k^{(j)}\} K \in S_c) \\
\text{Maximum attack: } \mathbf{V}_{max}^{(j)} &= \max(\{\mathbf{f}_k^{(j)}\} K \in S_c) \\
\text{Median attack: } \mathbf{V}_{median}^{(j)} &= \text{median}(\{\mathbf{f}_k^{(j)}\} K \in S_c) \\
\text{Minmax attack: } \mathbf{V}_{minmax}^{(j)} &= (\mathbf{V}_{min}^{(j)} + \mathbf{V}_{max}^{(j)}) \div 2 \\
\text{Modneg attack: } \mathbf{V}_{modneg}^{(j)} &= \mathbf{V}_{min}^{(j)} + \mathbf{V}_{max}^{(j)} - \mathbf{V}_{median}^{(j)} \\
\text{Randneg attack: } \mathbf{V}_{randneg}^{(j)} &= \begin{cases} \mathbf{V}_{min}^{(j)} & \text{with prob. } p \\ \mathbf{V}_{max}^{(j)} & \text{with prob. } 1 - p \end{cases}
\end{aligned} \tag{1}$$

Where $\mathbf{f}_K^{(j)}$ represents the j th component of the K th fingerprint. $\min(\cdot)$, $\max(\cdot)$, $\text{median}(\cdot)$ represent minimum function, maximum function and median function respectively.

It should be noted that the illegal users may collude with all the coefficients because they didn't know which coefficients are the components of the host vector. So they must compute the collusion copy with all the coefficients participating. We imitate formula (1) to construct all-component collusion model as follows.

$$\begin{aligned}
\text{Average attack: } I(x, y)_{aver} &= (\sum_{K \in S_c} I(x, y))/K \\
\text{Minimum attack: } I(x, y)_{min} &= \min((I(x, y))_{K \in S_c}) \\
\text{Maximum attack: } I(x, y)_{max} &= \max((I(x, y))_{K \in S_c}) \\
\text{Median attack: } I(x, y)_{median} &= \text{median}((I(x, y))_{K \in S_c}) \\
\text{Minmax attack: } I(x, y)_{minmax} &= (I(x, y)_{min} + I(x, y)_{max})/2 \\
\text{Modneg attack: } I(x, y)_{modneg} &= I(x, y)_{min} + I(x, y)_{max} - I(x, y)_{median} \\
\text{Randneg attack: } I(x, y)_{randneg} &= \begin{cases} I(x, y)_{min} & \text{with prob. } p \\ I(x, y)_{max} & \text{with prob. } 1 - p \end{cases}
\end{aligned} \tag{2}$$

Where $I(x, y)_k$ represents the coefficients of the k th image located at (x, y) . The rest parameters are same as formula (1). $I(x, y)_k$ may be frequency coefficient or pixel value.

3. Collusion attack analysis.

3.1. Collusion process. This paper suppose DCT domain as the embedding domain for most literatures embed fingerprints in DCT domain. The non-embedding domain represented with spatial domain and wavelet domain. Figures 1-3 show the collusion process.

Fig.1 shows the collusion process in DCT domain. The collusion operation is done after DCT operation. It is clear that the objects of collusion computing are DCT coefficients. Fig.2 shows the collusion process in spatial domain, and its operation objects are pixels. However, from the dashed frame in Fig.2 we know it is equivalent to a group of orthogonal transform are done before and after collusion operation on DCT coefficients. Similar with spatial domain collusion, Fig.3 is collusion in wavelet domain. There are two groups of transform before and after collusion operation on DCT coefficients. The outer contents of dashed frame are same in Figures 1-3. Their differences are manifested in the inner of

dashed frame. The collusion in non-embedding domain can be seen as a special collusion in embedding domain. The difference is that there are some orthogonal transforms before and after collusion operation.

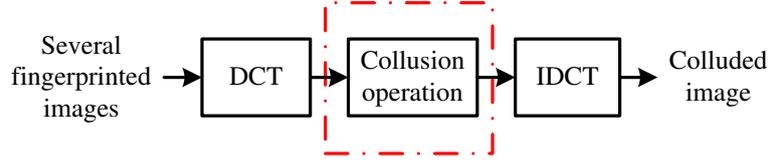


FIGURE 1. Collusion in DCT domain

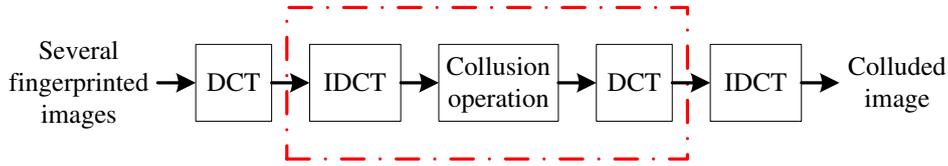


FIGURE 2. Collusion in spatial domain

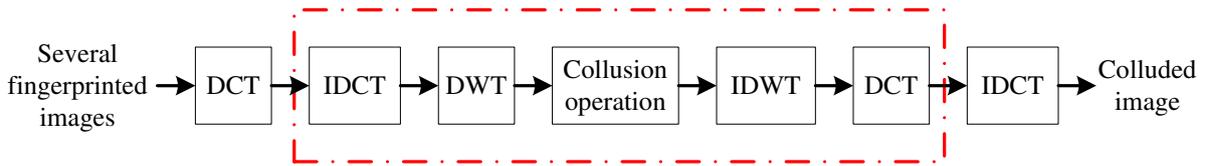


FIGURE 3. Collusion in wavelet domain

3.2. Theory analysis. In this section we analyze the effects of different collusion with all-components model. Suppose $\mathbf{M}_k(k \in \mathbf{S}_c)$ is the data to be operated, $g(\cdot)$ represents collusion operation. Then the collusion in DCT domain, spatial domain and wavelet domain are defined as follows.

$$Attack_{dct} = g(\mathbf{M}_k) \tag{3}$$

$$Attack_{spatial} = DCT(g(DCT^{-1}(\mathbf{M}_k))) \tag{4}$$

$$Attack_{wavelet} = DCT(DWT^{-1}(g(DWT(DCT^{-1}(\mathbf{M}_k)))))) \tag{5}$$

Where DCT , DCT^{-1} , DWT , DWT^{-1} denotes DCT, Inverse Discrete Cosine Transform(IDCT), Discrete Wavelet Transform(DWT) and Inverse Discrete Wavelet Transform(IDWT) respectively. In order to describe simply, we define the following notations.

- Attack X > Attack Y: Attack X is more effective than attack Y in defeating the fingerprinting.
- Attack X = Attack Y: Attack X and attack Y have the same performance in defeating the fingerprinting.
- Attack X \approx Attack Y: Attack X and attack Y have similar performance in defeating the fingerprinting.

The all-component collusion model is studied next from the colluders angle. For mean operation is linear computation, if $aver(\cdot)$ represents mean operation we can know

$$\begin{aligned} & DCT(aver(DCT^{-1}(\mathbf{M}_k))) \\ &= DCT(DCT^{-1}(aver(\mathbf{M}_k))) \\ &= aver(\mathbf{M}_k) \end{aligned} \quad (6)$$

$$\begin{aligned} & DCT(DWT^{-1}(aver(DWT(DCT^{-1}(\mathbf{M}_k)))))) \\ &= DCT(DWT^{-1}(DWT(DCT^{-1}(aver(\mathbf{M}_k)))))) \\ &= aver(\mathbf{M}_k) \end{aligned} \quad (7)$$

The formula (6) and (7) indicate that average attack in DCT domain, spatial domain and wavelet domain are same. Further we can obtain **conclusion 1**:

Attack DCT-aver=Attack Spatial-aver=Attack Wavelet-aver.

For maximum operation, minimum operation we have

$$\begin{aligned} & DCT(max(DCT^{-1}(\mathbf{M}_k))) \\ &< DCT(DCT^{-1}(max(\mathbf{M}_k))) \\ &= max(\mathbf{M}_k) \end{aligned} \quad (8)$$

$$\begin{aligned} & DCT(min(DCT^{-1}(\mathbf{M}_k))) \\ &< DCT(DCT^{-1}(min(\mathbf{M}_k))) \\ &= min(\mathbf{M}_k) \end{aligned} \quad (9)$$

$$\begin{aligned} & DCT(DWT^{-1}(max(DWT(DCT^{-1}(\mathbf{M}_k)))))) \\ &< DCT(max(DCT^{-1}(\mathbf{M}_k))) \\ &< DCT(DCT^{-1}(max(\mathbf{M}_k))) \\ &= max(\mathbf{M}_k) \end{aligned} \quad (10)$$

$$\begin{aligned} & DCT(DWT^{-1}(min(DWT(DCT^{-1}(\mathbf{M}_k)))))) \\ &< DCT(min(DCT^{-1}(\mathbf{M}_k))) \\ &< DCT(DCT^{-1}(min(\mathbf{M}_k))) \\ &= min(\mathbf{M}_k) \end{aligned} \quad (11)$$

Here $A < B$ indicates the variation amplitude of A is bigger than B . According to formula (2) we can get **conclusion 2**:

Attack DCT-min=Attack DCT-max>Attack Spatial-min=Attack Spatial-max>Attack Wavelet-min=Attack Wavelet-max.

Attack DCT-randneg>Attack Spatial-randneg>Attack Wavelet-randneg.

For median operation we have

$$DCT(median(DCT^{-1}(\mathbf{M}_k))) \approx median(\mathbf{M}_k) \quad (12)$$

$$DCT(DWT^{-1}(median(DWT(DCT^{-1}(\mathbf{M}_k)))))) \approx median(\mathbf{M}_k) \quad (13)$$

According to formula (2), (8), (9), **conclusion 3** can be obtained easily:

Attack DCT-mean \approx Attack Spatial-mean \approx Attack Wavelet-mean.

Attack DCT-minmax \approx Attack Spatial-minmax \approx Attack Wavelet-minmax.

Attack DCT-modneg \approx Attack Spatial-modneg \approx Attack Wavelet-modneg.

4. Experimental results. In order to examine the correctness of the conclusions, we employed 1000-dimension fingerprints which are random sequences following Gaussian distribution to test. 10 fingerprints were chosen randomly to be computed with the 7 kinds form in formula (2). The simulations are processed 10 groups in total. Figures 4-10 show the fingerprint histogram under each collusion. Fig.4 shows the results of averaging collusion. It is very obvious that the 3 figures in Fig.4 are same, which indicates that the results are consistent with conclusion 1. From Fig.5 we can know the descending order of fingerprints distribution deviation is Figs.5(a), 5(c) and 5(b). It means Fig.5(a) corresponds to the most seriously attack. The results prove the correctness of conclusion 2. The same results can be derived from Fig.6 and Fig.7. All the fingerprints distribution deviations for all the sub-images of Figures 8-10 are almost the same, so the three figures correspond to the attacks with similar intensity. The results are consistent with conclusion 3.

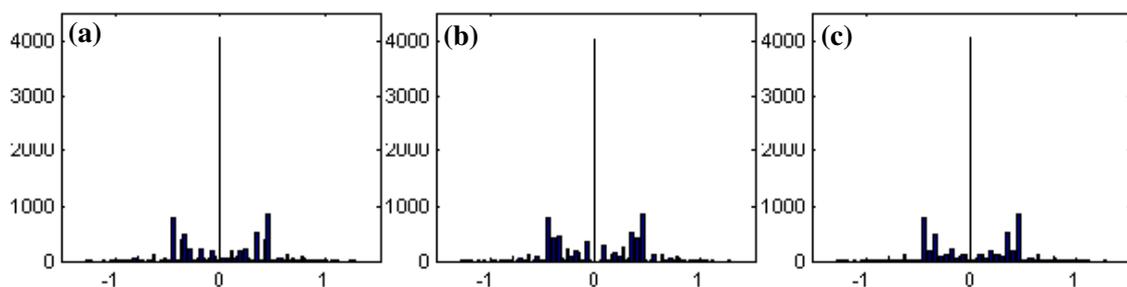


FIGURE 4. Fingerprint histogram under average attack (a)DCT (b)Spatial (c)Wavelet

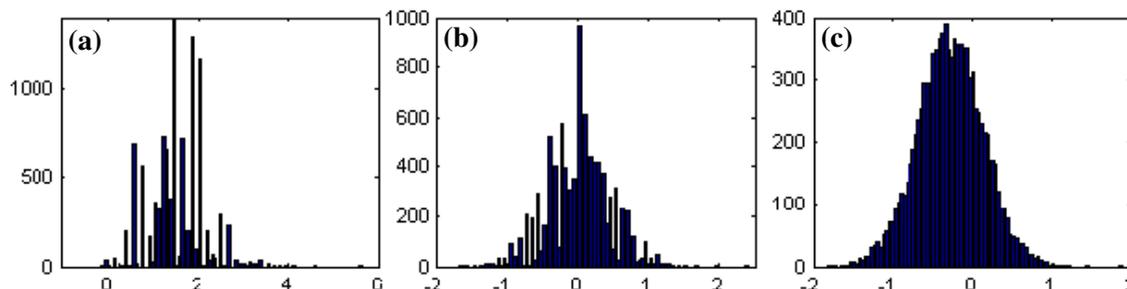


FIGURE 5. Fingerprint histogram under maximum attack (a)DCT (b)Spatial (c)Wavelet

In order to prove the correctness of conclusions further, an orthogonal fingerprint database which is composed of 1024 Gaussian sequences with 1024-dimension. Then 100 randomly selected fingerprints were embedded into 256×256 images with spread spectrum method [1,2]. The fingerprinted images were colluded with formula (1) and (2) and further compressed to get collusion copy. The colluders were traced according to the extracted fingerprint. Catching one colluder was studied in the paper [12], 100 groups of simulations were done with all kinds of collusion attacks. Fig.11 shows the results of the probabilities of detection.

There are six sub-images in Fig.11, and there are 4 curves in each sub-image. The DCT, DCT-all, Spatial and Wave curves represent the results of collusion with formula (1) in DCT domain, with formula (2) in DCT domain, with formula (2) in Spatial domain, with formula (2) in Wavelet domain, respectively.

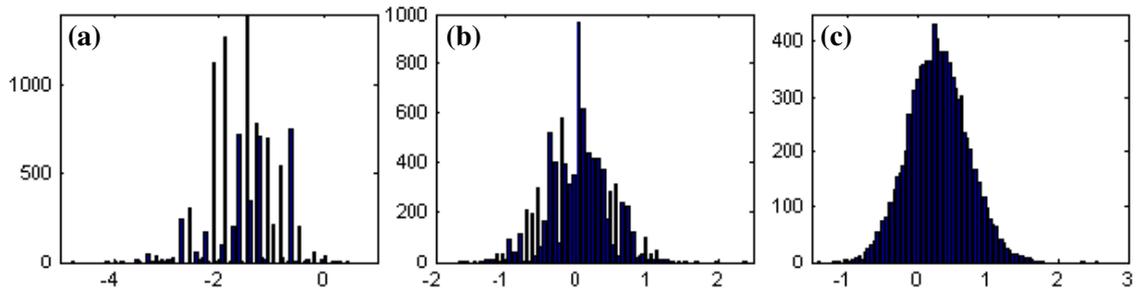


FIGURE 6. Fingerprint histogram under minimum attack (a)DCT (b)Spatial (c)Wavelet

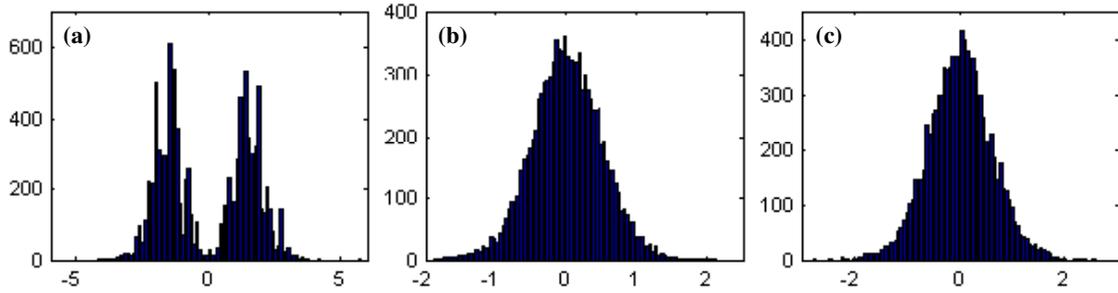


FIGURE 7. Fingerprint histogram under randneg attack (a)DCT (b)Spatial (c)Wavelet

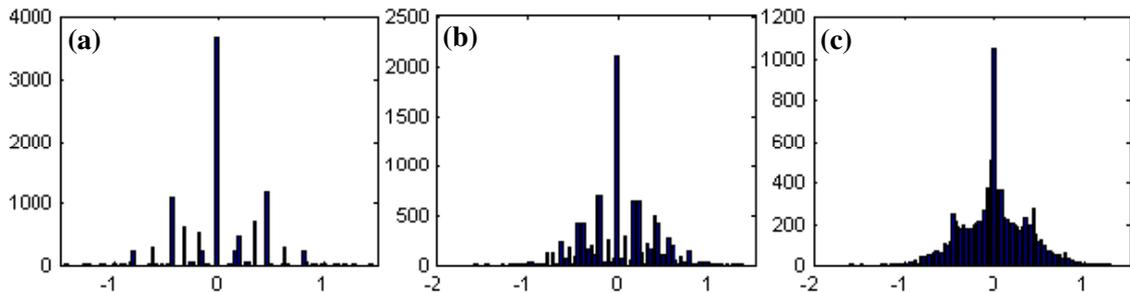


FIGURE 8. Fingerprint histogram under median attack (a)DCT (b)Spatial (c)Wavelet

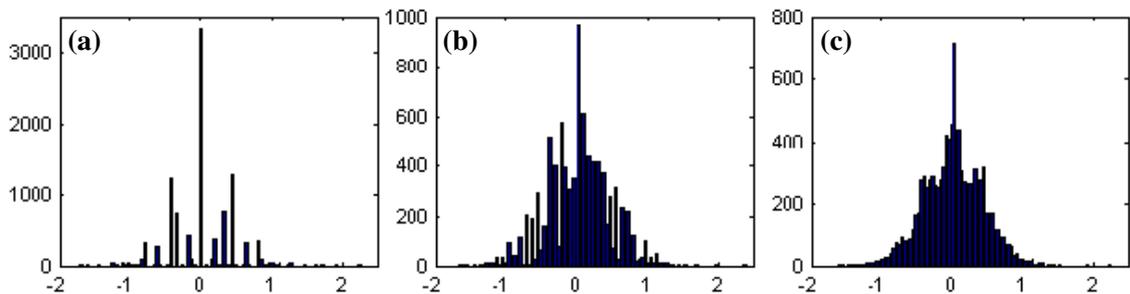


FIGURE 9. Fingerprint histogram under minmax attack (a)DCT (b)Spatial (c)Wavelet

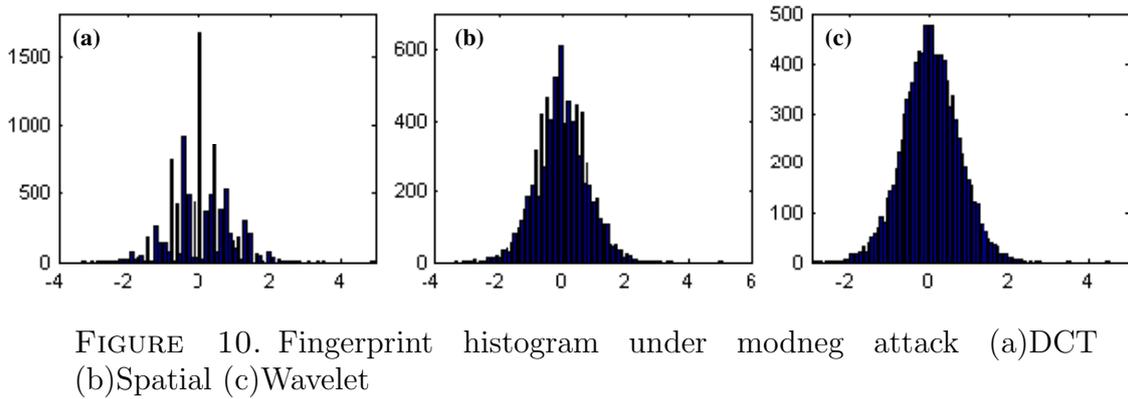


FIGURE 10. Fingerprint histogram under modneg attack (a)DCT (b)Spatial (c)Wavelet

Fig.11(a) shows the probabilities of detection under averaging attack. It is easy to know that all the curves in Fig.11(a) are coincident, which indicates the collusion copies are same with the four kinds of attacks and proves the conclusion 1. Figs.11(b) and 11(f) show the scenarios under maximum and randneg attack. The probabilities of detection under minimum attack is same as that maximum, so it is omitted for economization. From Figs.11(b) and 11(f) it is clear that the probabilities of detection under maximum attack with formula (1) in DCT domain is lowest, and the attack with formula (2) in DCT domain is close to it. The third is that with formula (2) in wavelet domain followed by that with formula (2) in spatial domain. The curves are consistent with conclusion 2 and Figures 5-7. Figs.11(c),11(d), and 11(e) are probabilities of detection under median attack, minmax attack and modneg attack. The curves in Figs.11(c),11(d), and 11(e) are close and they are also consistent with conclusion 3.

It can be conclude that the attack effect with partial-component collusion model in embedding domain is close to that with all-component collusion model, and the former slightly stronger than the latter. the attack effect with partial-component collusion model in embedding domain is stronger than that with all-component collusion model in non-embedding domain. So the proposed all-component model is more representative.

Generally, stronger attack often leads to bad quality of fingerprinted copy. The Peak-Signal-to-Noise Ratio(PSNR) is often used to measure the image quality. It is defined as follows

$$PSNR = 10 \log_{10} \left(\frac{255MN}{\sum_{x=1}^M \sum_{y=1}^N (I(x,y) - \hat{I}(x,y))^2} \right) \quad (14)$$

Where $I(x,y)$, $\hat{I}(x,y)$ denote the pixel located at (x,y) in original image and collusion copy respectively, M and N are their length and width.

The relationship between quality of colluded images and colluder number is shown in Fig.12 for realization comprehensively. The vertical axis denotes PSNR and the horizontal axis denotes colluder number. From Figs.12(a), 12(c), 12(d), and 12(e) we know PSNR increases with the increase of colluder number under average attack, median attack, min-max attack and modneg attack. It is because that in these attacks the coefficients will be close to their original value, and the greater number the closer it will be.

On the contrary, from Figs.12(b) and 12(f) PSNR decrease with the increase of colluder number when collude under minimum attack, maximum attack, randneg attack. Being similar with the former analysis, the coefficients deviate the original value when the colluded copies are obtained with these attacks. Naturally, more coefficient can make the deviation more severe.

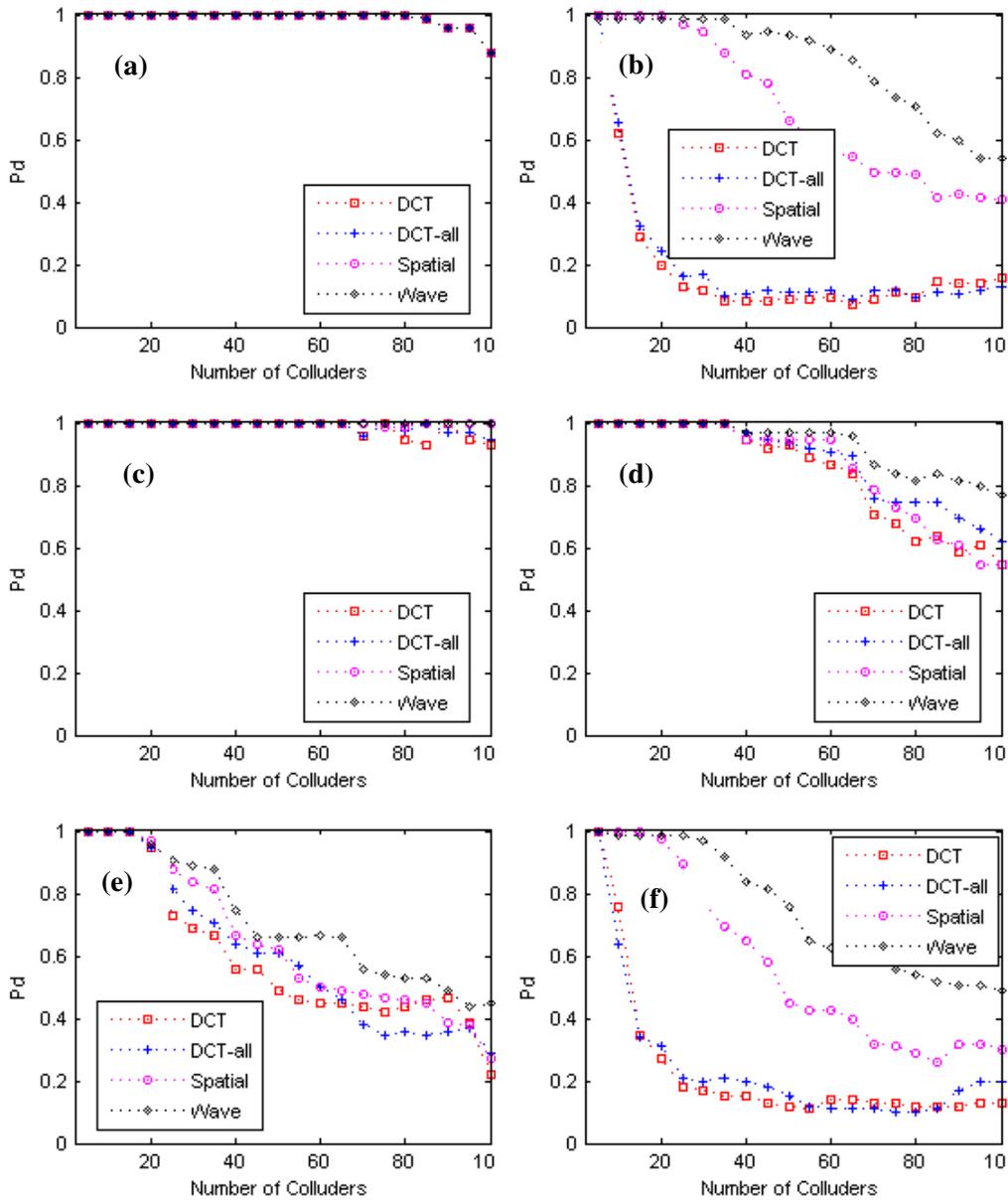


FIGURE 11. Probability of detection under collusion attacks (a) average (b) maximum (c) median (d) minmax (e) modneg (f) randneg

It can be concluded that from Fig.11 and Fig.12, when the quality of colluded image becomes well with the increase colluders it needs more colluders to defeat the fingerprinting. At the same time, when the quality of colluded image becomes bad with the increase colluders it needs less colluders to defeat. In fact it is in accord with the law that the more distorted image the more difficult to test information from it.

5. Conclusions. The prior works constructed collusion model in the embedding domain only for convenience. A new collusion model is constructed from the colluders' angle. We analyzed the difference of the two models in theory and obtained some conclusions. Large experimental results proves the conclusions. This can support the former model and can be a guide for designing fingerprinting system.

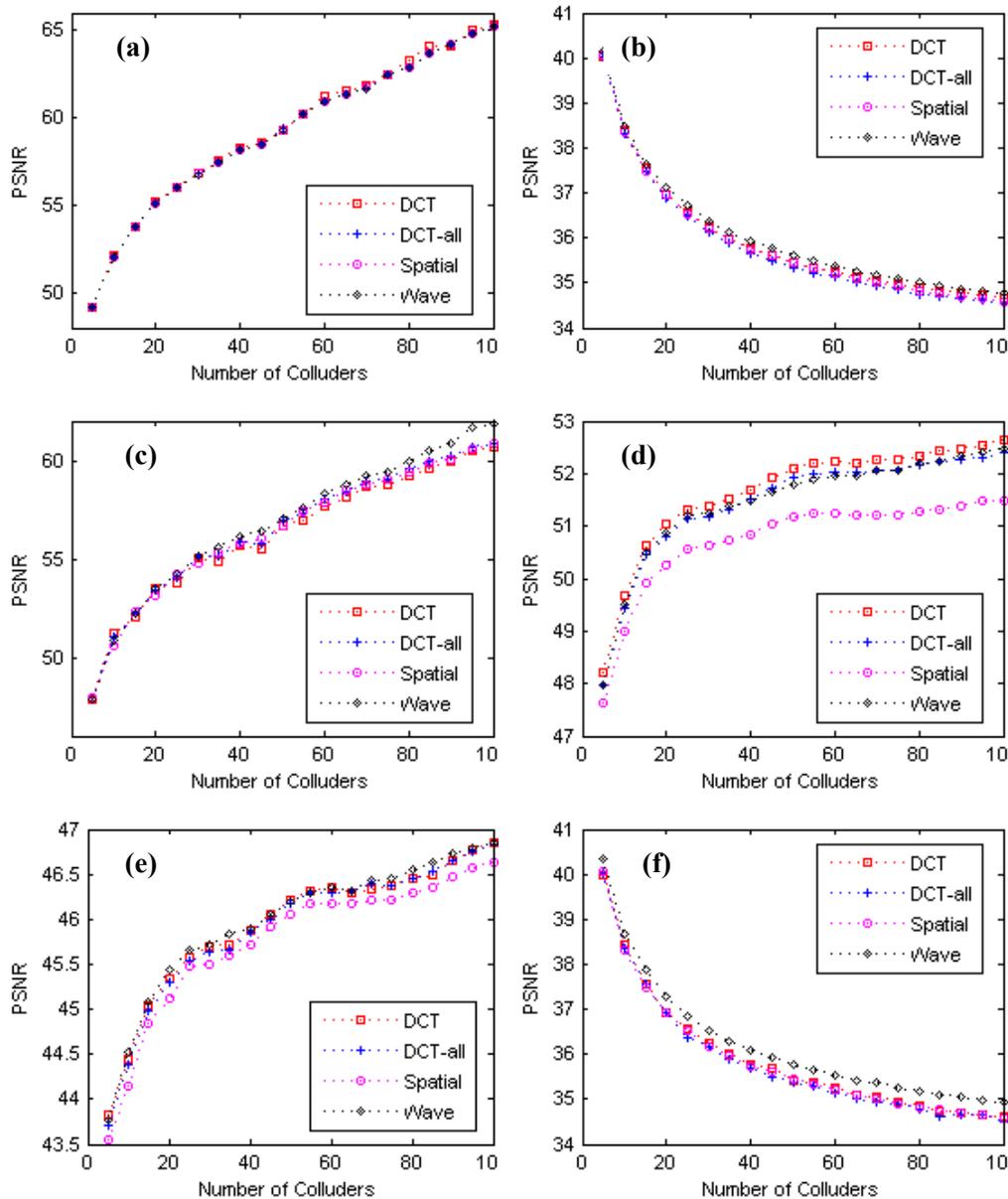


FIGURE 12. The relationship between PSNR and colluder number under collusion attacks (a) average (b) maximum (c) median (d) minmax (e) mod-neg (f) randneg

Acknowledgment. This work is supported by National Natural Science Foundation of China under Grant No.60802077, No.61003196, No.61105066 and No.60872136, the Fundamental Research Funds for the Central Universities under Grant No.K50510040007, China Postdoctoral Science Foundation Special Funded Project under Grant No.201104586 China Postdoctoral Science Foundation under Grant No.20100471415, Science Research Foundation of China University of Mining and Technology under Grant No.2009A022. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

- [1] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, Secure spread spectrum watermarking for multimedia, *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673-1687, 1997.

- [2] C. Podilchuk and W. Zeng, Image adaptive watermarking using visual models, *IEEE Journal on Selected Areas Communications*, vol. 16, no. 5, pp. 525-540, 1998.
- [3] H. C. Huang, Y. H. Chen, and A. Abraham, Optimized watermarking using swarm-based bacterial foraging, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 1, pp. 51-58, 2010.
- [4] H. Al-Qaheri, A. Mustafi, and S. Banerjee, Digital watermarking using ant colony optimization in fractional fourier domain, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 3, pp. 179-189, 2010.
- [5] D. Boneh and J. Shaw, Collusion-secure fingerprinting for digital data, *IEEE Trans. Information Theory*, vol. 44, no. 5, pp. 1897-1905, 1998.
- [6] A. Fiat and T. Tassa, Dynamic tracing traitors, *Proc. of Advances in Cryptology*, vol. 1666, pp. 354-371, 1999.
- [7] B. Pfizmann and M. Waidner, Asymmetric fingerprinting for larger collusions, *Proc. of the 4th ACM Conference Computer and Communication Security*, pp. 151-160, 1997.
- [8] F. Zane, Efficient watermark detection and collusion security, *Proc. of International Conference on Financial Cryptography*, vol. 1962, pp. 21-32, 2000.
- [9] W. Trappe, M. Wu, Z. J. Wang and K. J. R. Liu, Anti-collusion fingerprinting for multimedia, *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 1069-1087, 2003
- [10] X. W. Li, B. L. Guo and L. D. Li, A high capacity image fingerprinting scheme with blind detection, *Journal of Optoelectronicso Laser*, vol. 3, no. 3, pp. 446-450, 2011.
- [11] Z. J. Wang, M. Wu, H. V. Zhao, W. Trappe and K. J. R. Liu, Anti-collusion forensics of multimedia fingerprinting using orthogonal modulation, *IEEE Trans. Image Processing*, vol. 14, no. 6, pp. 804-821, 2005.
- [12] M. Wu, W. Trappe, Z. J. Wang and K. J. R. Liu, Collusion-resistant fingerprinting for multimedia, *IEEE Signal Processing Magazine*, pp. 15-27, 2004.
- [13] H. Gou and M. Wu, Data hiding in curves with application to fingerprinting Maps, *IEEE Trans. Signal Processing, Supplement on Secure Media*, vol. 53, no. 10, pp. 3988-4005, 2005.
- [14] S. He and M. Wu, Joint coding and embedding techniques for multimedia fingerprinting, *IEEE Trans. Information Forensics and Security*, vol. 1, no. 2, pp. 231-247, 2006.
- [15] A. L. Varna, S. He and M. Wu, Fingerprinting compressed multimedia signals, *IEEE Trans. Information Forensics and Security*, vol. 4, no. 3, pp. 330-345, 2009.
- [16] X. W. Li, B. L. Guo, F. J. Meng and L. D. Li, A novel fingerprinting algorithm with blind detection in dCT domain for images, *International Journal of Electronics and Communications*, vol. 65, no. 11, pp. 942-948, 2011.
- [17] H. Stone, Analysis of attacks on image watermarks with randomized coefficients, *NEC Research Institute, Technology Republication*, pp. 96-045, 1996.
- [18] H. V. Zhao, M. Wu, Z. J. Wang and K. J. R. Liu, Forensic analysis of nonlinear collusion attacks for multimedia fingerprinting, *IEEE Trans. Image Processing*, vol. 14, no. 5, pp. 646-661, 2005.
- [19] Z. Y. Chen, Z. Xiong, B. Cheng and L. Tang, Method against nonlinear collusion attack for digital fingerprinting, *Journal of Harbin Institute of Technology*, vol. 38, pp. 830-833, 2006.