SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Semi-Fragile Watermarking for Image Authentication based on Compressive Sensing

DU Ling[1,3], CAO XiaoChun[1,2*], ZHANG Wei[2], ZHANG XinPeng[4], LIU Na[2] & WEI JianGuo[1]

[1]*School of Computer Science and Technology, Tianjin University, Tianjin* 300072*, China;*
[2]*State Key Laboratory Of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100093*, China;*
[3]*School of Computer, Shenyang Aerospace University, Shenyang* 110136*,China;*
[4]*School of Communication and Information Engineering, Shanghai University, Shanghai* 200072*, China*

Appendix A Performance Comparison with Previous Approaches

The performance comparison of our proposed and previous approaches is shown in Table A1. Among these methods, multiple semi-fragile watermarks based methods such as in [1–5] are a type of traditional schemes. It can make the localization accurately based on authentication watermark, and content recovery tolerating some incidental manipulations. However, the recovery watermark is generated with redundancy to ensure its integrity under tamper happened. When the tampered areas are large, the extracted recovery watermark is not integrated, which will lead to *tampering/missing coincidence* problem. On the contrary, it will lead to *watermark-data waste* problem. Moreover, the methods based on Compressive Sensing (CS) such as in [6–8] are another type of classical methods, which can effectively solve above problems. However, the localization methods for tampered areas are almost based on block, which is not accurately. Moreover, it belongs to fragile algorithm and can be destroyed by incidental manipulations. Compared to the two currently class typical methods, we extend CS based image tamper localization and self-recovery method to semi-fragile. Although some other methods also adopt CS, their definitions only contribute to image tamper detection or identification, and mostly belong to fragile algorithms. In contrast, we specially consider the robustness of sparse signal recovery, and our algorithm has the ability to identify tamper regions at pixel-level, and recover the content modifications, even under incidental manipulations. We jointly consider the three aspects of image authentication: identification, localization and recovery. The proposed algorithm is able to answer these questions, i) Has the image been processed incidentally or maliciously? ii) Which part of the image has been processed? iii) What is the original content in the tampered area? Moreover, two applications are developed based on our technique, addressing the problems of content recovery and privacy protection.

Appendix B Evaluation on Color Images

Since color images have three RGB channels, the tamper localization and content recovery can be realized based on jointly consideration among different channels. For tamper localization, we use a meaningful binary image for authentication watermark generation, and each block is assigned for 4-bits watermark. Then, the authentication watermark is embedded into the three RGB channels respectively. By performing the corresponding extraction process, the corresponding 12-bits watermark can be extracted from each block, corresponding to 4-bits for each channel. For tamper localization, the final difference image identification are based on there 12-bits authentication watermark. Moreover, in some cases, the requirement for image recovery isn't color sensitive. Therefore, as for content recovery for color images, we adopt the gray-level image content for recovery watermark generation, which represents the image content sufficiently. The recovery watermarks are embedded into RGB channels respectively, and the corresponding content recovery are based on the jointly recovery watermark extraction. For performance analysis, we verify the color image authentication on database Kodak[1)], which contains 25 color images sized 512×768 or 768×512. The localization and recovery performance under different tamper ratios and incidental attacks on Kodak dataset are shown in Table B1 and Table B2.

* Corresponding author (email: caoxiaochun@iie.ac.cn)
1) http://www.r0k.us/graphics/kadak

Table A1 Comparison between different image authentication methods. At.1-At.4 stand for different attacks. At.1: JPEG compression, At.2: Noise addition, At.3: Brightness/Contrast adjustment, At.4: Format conversion. "•" indicates a "pass", "○" indicates a "fail", and "−" indicates there is no result provided by the author. √ indicates "can", and "×" indicates "can not".

Method	CS based	Localization	Recovery	Localization Robustness				Recovery Robustness			
				At.1	At.2	At.3	At.4	At.1	At.2	At.3	At.4
Ref [1]	×	pixel-level	√	•	−	−	•	•	−	−	•
Ref [2]	×	pixel-level	√	•	−	−	•	•	−	−	•
Ref [3]	×	pixel-level	√	•	•	−	•	−	−	−	−
Ref [4]	×	pixel-level	×	•	−	−	•				
Ref [5]	×	block-level	√	•	−	−	•	•	−	−	•
Ref [6]	√	pixel-level	×	−	•	•	−				
Ref [7]	√	pixel-level	×	•	−	•	•				
Ref [8]	√	block-level	√	○	○	○	○	○	○	○	○
Proposed	√	pixel-level	√	•	•	•	•	•	•	•	•

Appendix C Applications for Social Media

Nowadays, along with the development of social media, security and privacy are always two ever-green topics, due to the explosion of user generated data. Current social media suffers from two ways. In one way, the data is not protected and content can be modified for malicious purpose. In the other way, there is little privacy consideration for photo sharing.

Appendix C.1 Image Self-Recovery

With the advent of high-resolution digital cameras, powerful personal computers and sophisticated photo-editing software, the manipulation of photos is becoming more common. Figure C1 shows some examples of image content recovery for malicious tamper. The left three cases are real tamper examples. The first example is a News photo taken by a reporter, introduced that the man with passbook in his hands donated the money of his house demolition to help others. However, the passbook had been tampered with a box of medicines in a drug advertising for other purposes. The second case is also about false advertising. A glasses advertising posters is appeared at a street in Nanning, Guangxi. Parents accidentally found that the girl with a pair of glasses appeared in the advertising is actually forged from the poster of the Hope Project. The third case is about the star Stefanie Sun, who visited Africa, and measured the weight and height for the hungry children with the medical staff at a local health center. She put the baby dressed in a bag on the pounds uprightly. However, someone intentionally modifies this photo, removing the infant scale and accusing her of child abuse. In view of above situations, our proposed algorithm is an effective way for the protection of these social medias. In practice, we embed the watermark into the image before posting to social media websites. Once the original content is tampered and re-posted, we can authenticate and recover the image by inspecting the watermark. Note that our algorithm operates on pixel level authentication, such that the fake region has nowhere to hide.

Appendix C.2 Privacy Protection

In the recent years, there has been a rapid expansion of photo-sharing in social networks, but a major challenge has emerged in demonstrating that this may be at the price of individual privacy, including personally-identifiable information, such as face, license plates, door number, etc. Figure C2 shows some examples of privacy protection for license plates, person faces and door numbers. We embed the recovery watermark into the original image based on our method before they are posted. Owing to the purpose of privacy protection, the faces or license plates of corresponding images may be blurred, just as shown in the second column ((b), (e), (h)) of Figure C2. However, besides the publisher, other persons (such as trusted friends, colleagues, etc) who need to acquire the blurred content, it is necessary that they have the permissions for viewing the original content. In such cases, image recovery algorithm is of vital importance. The third column ((c), (f), (i)) of Figure C2 shows the recovery results by our method. It is worth noting that the social media websites only keep the the blurred photo, and the sensitive region is only visible to the people with the key to recover the content. This is much more secure, compared to maintaining the original copy on data servers and displaying differently to different users. In our case, the photo is secure even someone hacked the server.

References

1 Piva A , Bartolini F , Caldelli R. Self recovery authentication of images in the dwt domain. International Journal of Image and Graphics, 2005, 5(1): 149-166

2 Chamlawi R, Khan A , Idris A. Wavelet based image authentica- tion and recovery, Journal of Computer Science and Technology, 2007, 22(6):795-804 Nov. 2007.

3 Lv L, Fan H, Wang J, et al. A semi-fragile watermarking scheme for image tamper location and recovery. Journal of Theoretical and Applied Information Technology, 2012, 42(2): 287-291

Table C1 PDA/PFP for tamper localization with different tamper ratios for the Kodak image database.

Ratio+Attack	**kodim01**	**kodim02**	**kodim03**	**kodim04**	**kodim05**	**kodim06**	**kodim07**	**kodim08**
9.40%+Noise(rate=0.01)	1/0.04	0.99/0.05	0.99/0.05	0.99/0.09	0.99/0.03	0.99/0.37	1/0.05	1/0.04
13.62%+Intensity(gamma=1.2)	0.99/0.002	0.99/0.008	0.99/0.01	1/0.01	1/0.004	0.99/0.29	1/0.009	1/0.009
18.63%+JPEG(QF=90)	0.97/0.02	0.97/0.07	0.97/0.06	0.97/0.04	0.97/0.02	0.96/0.02	0.96/0.04	0.97/0.01
30.98%+Intensity(gamma=0.8)	0.98/0	0.98/0.002	0.99/0.002	0.98/0	0.98/0	0.98/0	0.98/0.001	0.98/0
Ratio+Attack	**kodim09**	**kodim10**	**kodim11**	**kodim12**	**kodim13**	**kodim14**	**kodim15**	**kodim16**
9.40%+Noise(rate=0.01)	1/0.1	1/0.09	0.99/0.06	0.99/0.24	0.99/0.01	1/0.04	0.97/0.40	1/0.04
13.62%+Intensity(gamma=1.2)	1/0.004	1/0.007	0.98/0.005	0.99/0.004	0.99/0.004	1/0.004	0.98/0.33	1/0.004
18.63%+JPEG(QF=90)	0.96/0.12	0.96/0.04	0.97/0.02	0.96/0.04	0.97/0.02	0.96/0.04	0.97/0.04	0.96/0.02
30.98%+Intensity(gamma=0.8)	0.99/0	0.99/0.002	0.98/0.003	0.99/0.001	0.98/0	0.98/0	0.98/0.003	0.99/0.002
Ratio+Attack	**kodim17**	**kodim18**	**kodim19**	**kodim20**	**kodim21**	**kodim22**	**kodim23**	**kodim24**
9.40%+Noise(rate=0.01)	0.99/0.04	0.98/0.06	1/0.09	1/0.42	1/0.03	1/0.08	0.99/0.05	0.99/0.06
13.62%+Intensity(gamma=1.2)	0.99/0.005	0.99/0.005	1/0.005	1/0.55	1/0.004	1/0.004	1/0.02	1/0.02
18.63%+JPEG(QF=90)	0.97/0.04	0.97/0.02	0.97/0.03	0.96/0.18	0.96/0.04	0.96/0.05	0.96/0.02	0.96/0.05
30.98%+Intensity(gamma=0.8)	0.98/0.002	0.98/0	0.98/0.002	0.99/0	0.99/0.002	0.99/0	0.99/0.002	0.99/0.002

Table C2 PSNR/VIF for tamper recovery with different tamper ratios for the Kodak image database.

Ratio+Attack	**kodim01**	**kodim02**	**kodim03**	**kodim04**	**kodim05**	**kodim06**	**kodim07**	**kodim08**
3.16 %	37.76/0.76	41.89/0.86	40.17/0.88	40.58/0.88	33.12/0.78	34.92/0.78	40.45/0.87	31.48/0.72
9.40%+Intensity(gamma=0.9)	28.84/0.72	27.40/0.77	28.61/0.79	28.81/0.80	26.63/0.73	27.58/0.72	29.01/0.81	27.22/0.68
13.62%+Noise(rate=0.0001)	28.13/0.72	28.95/0.77	29.72/0.80	28.88/0.82	27.44/0.75	26.80/0.71	28.39/0.80	26.50/0.68
18.63%+ bmp→tif	35.41/0.76	39.36/0.85	36.82/0.84	38.21/0.86	30.75/0.76	30.21/0.72	37.87/0.87	29.04/0.69
Ratio+Attack	**kodim09**	**kodim10**	**kodim11**	**kodim12**	**kodim13**	**kodim14**	**kodim15**	**kodim16**
3.16 %	40.20/0.83	39.10/0.86	38.22/0.81	39.83/0.86	33.26/0.76	36.90/0.83	29.76/0.82	40.74/0.83
9.40%+Intensity(gamma=0.9)	29.21/0.74	29.11/0.78	28.17/0.74	28.60/0.76	27.83/0.71	28.02/0.77	25.20/0.75	28.55/0.74
13.62%+Noise(rate=0.0001)	28.67/0.76	28.95/0.80	29.02/0.76	27.55/0.76	27.22/0.71	28.17/0.77	25.40/0.73	27.72/0.76
18.63%+ bmp→tif	36.07/0.80	35.93/0.83	36.46/0.83	35.08/0.82	30.69/0.73	34.01/0.81	26.38/0.71	36.21/0.81
Ratio+Attack	**kodim17**	**kodim18**	**kodim19**	**kodim20**	**kodim21**	**kodim22**	**kodim23**	**kodim24**
3.16 %	36.30/0.86	35.63/0.82	37.74/0.81	19.87/0.77	37.52/0.79	38.99/0.84	36.77/0.87	31.68/0.81
9.40%+Intensity(gamma=0.9)	26.59/0.79	26.34/0.76	28.05/0.70	15.37/0.68	29.20/0.73	29.16/0.76	28.70/0.79	26.98/0.75
13.62%+Noise(rate=0.0001)	27.75/0.78	28.73/0.77	29.03/0.75	11.81/0.64	29.11/0.75	29.17/0.78	28.58/0.78	25.75/0.73
18.63%+ bmp→tif	32.53/0.80	35.32/0.82	35.71/0.80	18.91/0.65	35.50/0.79	37.03/0.83	34.95/0.83	28.53/0.73

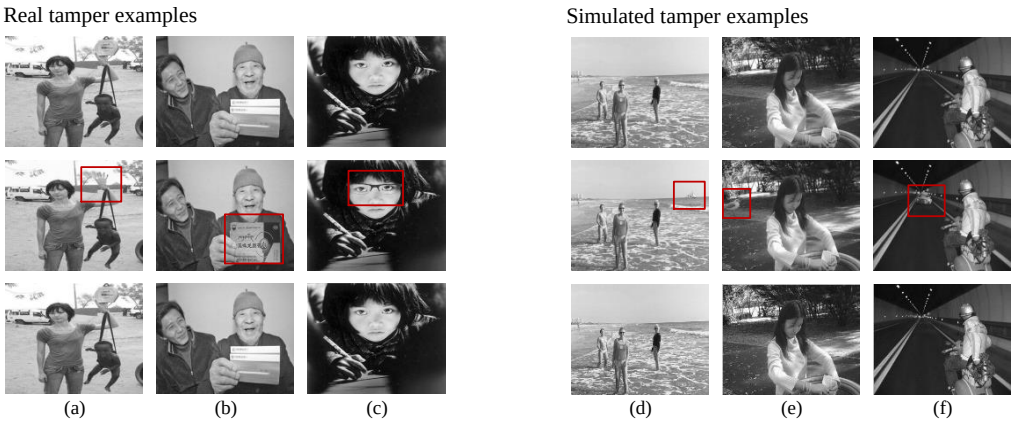

Figure C1 Example results of image self-recovery for image forgery. The first row shows original image. The second row shows the corresponding tampered images. The third row are the recovery results by our method.

4 Wu X , Hu J , Gu Z , et al. A secure semi-fragile watermark- ing for image authentication based on integer wavelet transform with parameters. Proceedings of the 2005 Australasian workshop on Grid computing and e-research, Australia , 2005. 75-80

5 Lin C, Chang S. Semi-fragile watermarking for authenticating JPEG visual content. Proceedings of SPIE, Security and Watermarking of Multimedia Content, California, 2000. 140-151

6 Tagliasacchi M , Valenzise G , Tubaro S et al. A compressive sensing based watermarking scheme for sparse image tampering identification. IEEE International Conference on Image Processing, Cairo, 2009. 1265-1268

7 Tagliasacchi M , Valenzise G , Tubaro S. Hash-based identification of sparse image tampering. IEEE Transactions on Image Processing, 2009, 18(11):2491-2504

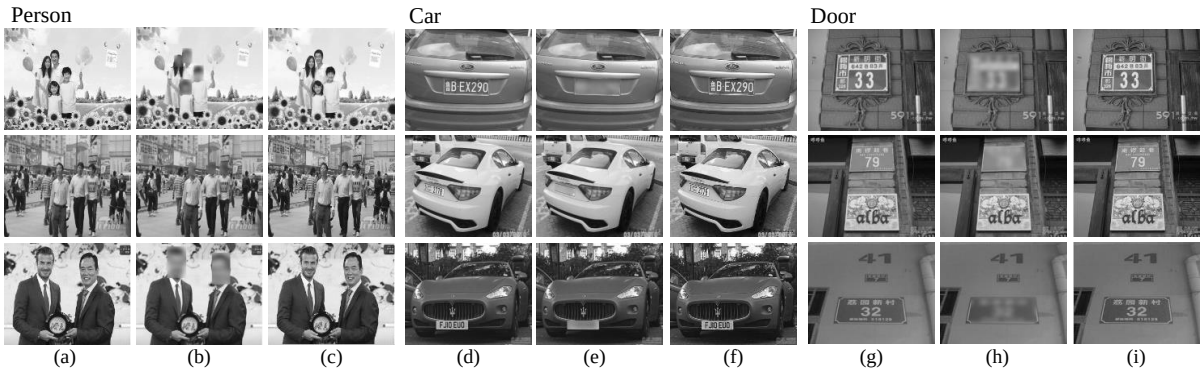

Figure C2 Examples of privacy protection for faces ((a), (b), (c)), car plates ((d), (e), (f)), and door tag((g), (h), (i)). (a), (d), (g) are original image, (b), (e), (h) are the corresponding images by blurring for the purpose of privacy protection. (c), (f), (i) are recovery results by our method. Note that the images are all taken from social media websites.

8 Zhang X , Wang S , Qian Z , et al. Watermarking with flexible self-recovery quality based on compressive sensing and compositive reconstruction. IEEE Transactions on Information Forensics and Security, 2011 6(4):1223-1232