SCIENCE CHINA
Information Sciences

• LETTER •

May 2016, Vol. 59 059104:1–059104:3
doi: 10.1007/s11432-016-5542-8

Semi-fragile watermarking for image authentication based on compressive sensing

Ling DU[1,3], Xiaochun CAO[1,2*], Wei ZHANG[2], Xinpeng ZHANG[4], Na LIU[2] & Jianguo WEI[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*

Received September 27, 2015; accepted November 24, 2015; published online April 8, 2016

Citation Du L, Cao X C, Zhang W, et al. Semi-fragile watermarking for image authentication based on compressive sensing. Sci China Inf Sci, 2016, 59(5): 059104, doi: 10.1007/s11432-016-5542-8

Dear editor,

With the aid of sophisticated photoediting software, multimedia content security is becoming increasingly prominent [1, 2]. The problems of verifying the integrity and recovering the alterations have been a major research topic for information security and multimedia [3–6]. Semi-fragile watermarking techniques are more powerful in dealing with such cases, which are fragile to malicious modifications meanwhile robust to incidental content-preserving manipulations. Compared to traditional replication based methods [7, 8], compressive sensing (CS) [9] based methods can effectively solve tampering/missing coincidence or watermark-data waste problem, and gain great success for image authentication.

In this letter, we extend CS based image tamper localization and self-recovery method [10] to semi-fragility. As shown in Figure 1(a), two watermarks, which can be regarded as image integrity and original content representations, are adopted for authentication and recovery. Although some other methods also adopt CS, their definitions only contribute to tamper detection or identification, and mostly belong to fragile algorithms. Unlike conventional CS-based approaches, we jointly consider three aspects of image authentication: identification, localization and recovery. Our algorithm is able to localize the tampered region at pixel level, as well as recover the original content.

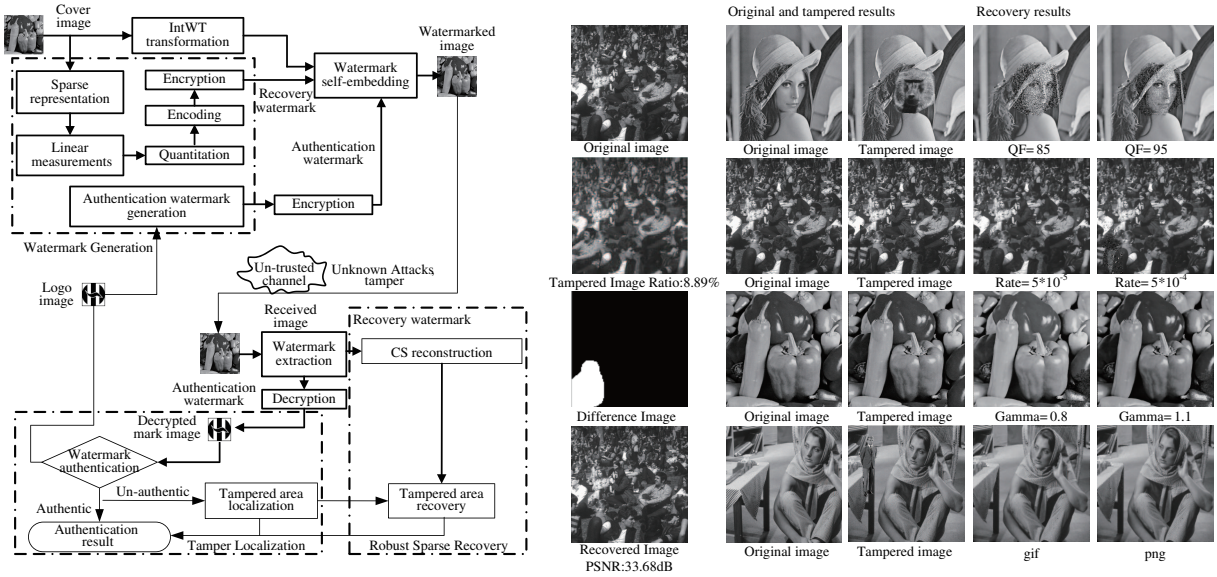
Watermark generation and embedding. Assume cover image $\boldsymbol{I}(N_1 \times N_2)$, and both N_1 and N_2 are multiples of 8, $N_s = N_1 \times N_2$. Here, we divide it into nonoverlapping 8×8 blocks. For tamper localization, we use a gray-level image $\boldsymbol{W}$ $(N_1/4 \times N_2/4)$, which is a concatenation of some meaningful logos, for authentication watermark generation. For security purpose, $\boldsymbol{W}$ is encrypted to $\boldsymbol{W}_A$ based on a chaotic function before embedding. Divide $\boldsymbol{W}_A$ into $N_s/64$ sets according to block number B $(N_s/64)$. Therefore, each block is assigned for 4 values (32 bits). Denote the assigned vector for block ka as $\boldsymbol{w}_{ka}(ka = 1, 2, ..., B)$, take it as the final authentication watermark.

For recovery watermark generation, we first apply one-level IntWT to the original image and obtain the low frequency subband LL_1 (with size $N_1/2 \times N_2/2$). For LL_1, segment it into nonover-

* Corresponding author (email: caoxiaochun@iie.ac.cn)

The authors declare that they have no conflict of interest.

© Science China Press and Springer-Verlag Berlin Heidelberg 2016



(a) Framework (b) Tamper localization and recovery (c) Recovery results (tamper+ incidental attacks)

Figure 1 Sketch of the proposed image authentication scheme and results.

lapping 4×4 blocks, and next, for each block, perform discrete cosine transform (DCT) and form the corresponding DCT coefficients into a vector $\boldsymbol{c}_k = [c_k(1), c_k(2), ..., c_k(16)]$ after zigzag reordering, where $1 \leqslant k \leqslant N_s/64$. Then, a pseudo-random permutation for all the blocks is performed and divide blocks into $G(N_s/4096)$ groups, each of which consists of 64 blocks. Therefore, the coefficients for each group form a new vector $\boldsymbol{v}_i = [\boldsymbol{c}_{k_1}, \boldsymbol{c}_{k_2}, ..., \boldsymbol{c}_{k_{64}}]^{\mathrm{T}}$, where $i = 1, 2, ..., G$. For each group, the reference vector $\boldsymbol{r}_i$ is calculated as $\boldsymbol{r}_i = \boldsymbol{\Phi}\boldsymbol{v}_i$, where $\boldsymbol{\Phi}_{M'\times N'}$ is a pseudo-random matrix drawn from the Gaussian distribution $N(0, 1/N')$ generated from a secret key known only to the embedder and decoder. Here, M' and N' are set as 768 and 1024 respectively. Denote the reference matrix as $\boldsymbol{R} = [\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_G]$. Then, make a permutation for $\boldsymbol{R}$ into $\boldsymbol{R}_T$. After that, we divide $\boldsymbol{R}_T$ into $N_s/64$ sets and each block is assigned for 12 reference values. Denote the reference vector for block kr as $\boldsymbol{w}_{kr}(kr = 1, 2, ..., B)$, and take it as the final recovery watermark.

To obtain better imperceptibility, for watermarks $\boldsymbol{w}_{ka}$ and $\boldsymbol{w}_{kr}(ka, kr = 1, 2, ..., N_s/64)$, the preprocessing process is performed by $\boldsymbol{w}_{ka}^{\mathrm{int}} = \mathrm{floor}(\boldsymbol{w}_{ka} - \mu_a)/\varepsilon_a$ and $\boldsymbol{w}_{ka}^{\mathrm{mod}} = \mathrm{mod}(\boldsymbol{w}_{ka} - \mu_a, \varepsilon_a)$, where $\boldsymbol{w}_{ka}^{\mathrm{int}}$ and $\boldsymbol{w}_{ka}^{\mathrm{mod}}$ are the modified authentication watermarks. Similarly, we get $\boldsymbol{w}_{kr}^{\mathrm{int}}$ and $\boldsymbol{w}_{kr}^{\mathrm{mod}}$ for recovery watermark with μ_r and ε_r. Here, μ_a and μ_r are mean values. For authentication watermark, as it is derived from gray-level logo image, μ_a is set to 128. As the matrix $\boldsymbol{\Phi}$ for recovery watermark generation is Gaussian distribution with zero mean, the reference values also approximately meet Gaussian distributions with zero mean, according to the central limit theorem. Therefore, we set μ_r=0, ε_a =ε_r = 8.

For watermark embedding, we segment image $\boldsymbol{I}$ into nonoverlapping 8×8 blocks and the number of blocks is $N_s/64$. After that, one level IntWT for each block is computed. Denote the horizontal and vertical wavelet subbands as HL_{k1} and LH_{k1} for block k. The two subbands are chosen for embedding to grant a good tradeoff between imperceptibility and robustness.

Tamper localization and self-recovery. For tamper localization, the difference image $\boldsymbol{D}$ between extracted authentication watermark image $\boldsymbol{W}^{\mathrm{e}}$ and original authentication image $\boldsymbol{W}^{\mathrm{o}}$ (generated in the same manner) is computed firstly by

$$\boldsymbol{D}_{ij} = \begin{cases} 1, \text{ if } \quad |\boldsymbol{W}_{ij}^{\mathrm{o}} - \boldsymbol{W}_{ij}^{\mathrm{e}}| > \delta, \\ 0, \text{ otherwise}, \end{cases} \tag{1}$$

where $1 \leqslant i \leqslant N_1/4,\ 1 \leqslant j \leqslant N_2/4$. $\boldsymbol{D}_{ij} = 1$ (white pixel in $\boldsymbol{D}$) means extraction error (tampered area). When the image is suffered from malicious modifications, most error pixels will gather together with high probability. As for incidental manipulations, the pixels with watermark error will be isolated. Therefore, we further handle such misdeclaration, including false alarm and false dismissal. Finally, the matrix $\boldsymbol{D}$ is adjusted to $\boldsymbol{D}_f$. The positions corresponding to the white pixels in $\boldsymbol{D}_f$ are taken as the tampered area.

For content recovery, after recovery watermark $\boldsymbol{w}_{kr}$ is acquired, we apply one-level IntWT to get LL_1 subband of the received image, and make division for block and group. By tamper localiza-

tion, we obtain the corresponding tampered and reserved blocks. For a certain group of LL_1 subband, assume there are z reserved blocks among the 64 ones, and r reserved blocks of the received image from which the reference values for this group can be extracted. That is, the number of correctly calculated coefficients N_β for the group is $16 \times z$, and the number of reference values M_α extracted from the reserved blocks is $12 \times r$. Denote $\boldsymbol{\Phi}^{'}$ with size $M_\alpha \times N'$ as a matrix with rows taken from $\boldsymbol{\Phi}_{M' \times N'}$ corresponding to the extractable reference values M_α. Therefore, there exists $\boldsymbol{r'} = \boldsymbol{\Phi}^{'}\boldsymbol{v}$, where $\boldsymbol{r'} \in \mathbb{R}^{M_\alpha \times 1}$. For the coefficient vector $\boldsymbol{v}$, there are N_β coefficients that can be calculated from z reserved blocks, and they form a new vector $\boldsymbol{v}_R = [\boldsymbol{c}_{k_1}, \boldsymbol{c}_{k_2}, ..., \boldsymbol{c}_{k_z}]^{\mathrm{T}}$. Moreover, there are N_γ coefficients calculated from $64 - z$ tampered blocks, which form vector $\boldsymbol{v}_T = [c_{k_1}, c_{k_2}, ..., c_{k_{64-z}}]^{\mathrm{T}}$. Therefore, $\boldsymbol{r'} = \boldsymbol{\Phi}^{'}\boldsymbol{v}$ can be rewritten as $\boldsymbol{r'} = \boldsymbol{\Phi}_R\boldsymbol{v}_R + \boldsymbol{\Phi}_T\boldsymbol{v}_T$, where $\boldsymbol{\Phi}_R$ and $\boldsymbol{\Phi}_T$ are matrices whose columns are derived from $\boldsymbol{\Phi}^{'}$ corresponding to the coefficient values in $\boldsymbol{v}_R$ and $\boldsymbol{v}_T$. Denote

$$\boldsymbol{u} = \boldsymbol{r'} - \boldsymbol{\Phi}_R\boldsymbol{v}_R = \boldsymbol{\Phi}_T\boldsymbol{v}_T, \tag{2}$$

where $\boldsymbol{u} \in \mathbb{R}^{M_\alpha \times 1}$, the size of $\boldsymbol{\Phi}_T$ is $M\alpha \times N_\beta$.

As the M_α reference and N_β coefficient values can be obtained from the reserved blocks, content recovery is dependent on the $\boldsymbol{v}_T$ calculation. As $\boldsymbol{v}_T$ is from the DCT coefficients, most values in $\boldsymbol{v}_T$ are close to zero. That is, $\boldsymbol{v}_T$ is sparse. Therefore, if $M\alpha$ is less than N_β, we can approximately reconstruct $\boldsymbol{v}_T$ based on CS theory. Otherwise, if $M\alpha$ is more than N_β, $\boldsymbol{v}_T$ can be resolved by $\boldsymbol{v}_T = [\boldsymbol{\Phi}_T]^{-1}\boldsymbol{u}$, where $[\boldsymbol{\Phi}_T]^{-1}$ is the pseudo-inverse of $\boldsymbol{\Phi}_T$. After combining $\boldsymbol{v}_T$ with $\boldsymbol{v}_R$, the coefficient vector for this group is retrieved. Finally, the LL_1 subband and the corresponding approximate cover image are reconstructed.

Experiment. In our experiment, parameter δ is set as 4, which is an empirical value. The PSNRs of the watermarked image are almost more than 30 dB. Figure 1(b) shows an example of tamper area localization and recovery. From the difference image, we can see that the tampered area is perfectly located by identifying the area concentrated by error pixels. Compared to the block-based methods, the proposed scheme can locate the tampered area at pixel-level. For the robustness analysis, Figure 1(c) shows some examples of recovery results under malicious tamper with respect to different tamper ratios and some incidental modifications (JPEG compression, slight noise addition, brightness/contract adjustment and format conversion). Most values of visual information fidelity (VIF) for recovery are more than 0.8, which are acceptable for the visual quality. Moreover, two applications are developed for photo sharing in current social media environment.

Conclusion. This work inherits the merits of robust image content self-recovery for incidental manipulations and flexible recovery quality for malicious modifications, and targets the image authentication issue based on semi-fragile digital watermarking and CS. Unlike conventional CS-based approaches, which mainly focus on image tamper detection or identification, our algorithm can identify tamper regions with pixel-level accuracy and recover the malicious/incidental modifications.

Acknowledgements This work was supported by National High Technology Research and Development Program of China (Grant No. 2014BAK11B03), and National Basic Research Program of China (Grant No. 2013CB329305), National Natural Science Foundation of China (Grant Nos. 61422213, 61170185), and "Strategic Priority Research Program" of the Chinese Academy of Sciences (Grant No. XDA06010701).

References

1 Li J, Li X L, Yang B, et al. Segmentation-based image copy-move forgery detection scheme. IEEE Trans Inf Foren Secur, 2015, 10: 507–518
2 Ren Y J, Shen J, Zheng Y H, et al. Efficient data integrity auditing for storage security in mobile health cloud. Peer-to-Peer Netw Appl, in press. doi: 10.1007/s12083-015-0346-y
3 Guo P, Wang J, Li B, et al. A variable threshold-value authentication architecture for wireless mesh networks. J Internet Tech, 2014, 15: 929–936
4 Xia Z H, Wang X H, Sun X M, et al. Steganalysis of LSB matching using differencesbetween nonadjacent pixels. Multimedia Tools Appl, in press. doi: 10.1007/s11042-014-2381-8
5 Pan Z Q, Zhang Y, Kwong S. Efficient motion and disparity estimation optimization for low complexity multiview video coding. IEEE Trans Broadcast, 2015, 61: 166–176
6 Liu S, Song Z, Liu G, et al. Street-to-shop: cross-scenario clothing retrieval via parts alignment and auxiliary set. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, 2012. 3330–3337
7 Chamlawi R, Khan A, Idris A. Wavelet based image authentication and recovery. J Comput Sci Tech, 2007, 22: 795–804
8 Lv L, Fan H, Wang J, et al. A semi-fragile watermarking scheme for image tamper location and recovery. J Theor Appl Inf Tech, 2012, 42: 287–291
9 Zhao G H, Shen F F, Wang Z Y, et al. A high quality image reconstruction method based on nonconvex decoding. Sci China Inf Sci, 2013, 56: 112103
10 Li Y, Du L. Semi-fragile watermarking for image tamper localization and self-recovery. In: Proceedings of International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Wuhan, 2014. 328–333