Method for Detection of an Image Tampering and Partial Recovery

David, Asatryan
Institute for Informatics and Automation Problems
Yerevan, Armenia
e-mail: dasat@ipia.sci.am

Naira, Asatryan
Center for Critical Technologies
Russian-Armenian (Slavonic) University
Yerevan, Armenia
e-mail: naira1973@yandex.ru

Natalya, Lanina
Research Center for Critical Technologies
Russian-Armenian (Slavonic) University
Yerevan, Armenia
e-mail: NSL@RAU.am

Alexandr, Petrosyan
Research Center for Critical Technologies
Russian-Armenian (Slavonic) University
Yerevan, Armenia
e-mail: petrosalex@mail.ru

ABSTRACT
In this paper, the problems of detection of the fact of unauthorized changing the content of a digital image, determination of distorted parts and partial recovering of the content from the damaged parts, are considered. Most of types of image tampering is based on the well known operations of Copy-Paste, Copy-Move or of its combinations. In many investigations the basic approach for tamper detection is based on creating a watermarking procedure in such a way that the type and parameters of distortion of the watermark from an attack allow to make an inference on the presence of a fraud, to locate the damaged parts of the image and even to recover them. In this paper a watermarking algorithm is used which allows the manipulations over the images at different ratios of sizes of image-container and watermark. Different schemes of protection from a fraud and recovery of distorted parts of an image are suggested. By means of numerical experiments the effectiveness of the suggested schemes and algorithms is shown.

Keywords
Image tampering, watermark compressing, authentication, integrity, recovery, combined approach

1. INTRODUCTION
With the emergence and development of image protection technology by means of embedding into it a digital watermark [1], the problem of establishing the authenticity and integrity of the image has aroused the interest of researchers. Variety of approaches and schemes for detection of the unauthorized access and distortion of protected information was suggested. With the assumption that tampering will alter also a watermark, an image can be authenticated by verifying that the extracted watermark is the same as that which was inserted or it is similar to that.

The most common type of fraud is an attack to the content of an image. In the literature it is named Copy-Paste, Copy-Move or Collage attack, when one of the parts of the image is copied to another part of the same or a different image, thereby altering the original content. In [2] several types of similar attacks are examined.

In practice, it is often sufficient to establish the fact of falsification or counterfeiting of an image. However, some watermarking algorithms allow also the identifying areas of damaged image [3, 4]. For example, in [4] a binary watermarking algorithm is offered, an example of a radiograph tampering is shown and a method for detecting the fact of tampering and locations of distortion over the image are given.

In the literature there are also proposed watermarking algorithms that not only detect the distortion of the image, but also are restoring the original information. One of the restoring methods is the using of fragments of the image-container instead of the watermark. If watermarking algorithm is resistant to various kinds of attacks, then the extracted watermark, even if it was distorted, might contain some information on location or configuration of damaged part of the image. Thus, it is possible to recover portions of the image that have been cropped out, replaced, damaged, or otherwise tampered, without accessing the original image [5, 6, and 7].

It is necessary to note that for self-embedding an image into itself one needs a universal enough watermarking algorithm, which allows manipulations with images of arbitrary types and arbitrary sizes. One of the approaches to create such algorithms is to combine the space-domain and frequency-domain techniques [8]. Really, using preliminary compression of a watermark in the frequency domain one can increase the volume of embedded information at admissible lost of quality.

The approach taken in [8] and related combined algorithm has been studied in detail in [9], resulting was greatly expanded their area of application.

In this paper, a method of application of combined algorithm to solve the problems of detecting the fact of an image

tampering, to identify the distorted parts and to recovery them subsequently, is proposed.

2. METHOD

We give a brief description of the relevant procedures, limiting to considering only the Gray Scale images.

2.1 Watermark embedding algorithm

2.1.1. Preliminary processing of the watermark

Before the embedding the watermark into the image-container it is transformed as follows:

- Watermark $W = w_{kl}$ is splitted into blocks of size $b \times b$ pixels, where b can take the values 4, 8 or 16 depending on the size of watermark-image, $l = 0,1,...,L-1$; $k = 0,1,...,K-1$;
- matrix of coefficients of the Discrete Cosine Transform (DCT) is calculated for each block, $DCT = (DCT_{ij})$, $i, j = 0,1,...,b-1$;
- $t \times t$ DCT coefficients of each block are stored, the rest is discarded;
- stored coefficients of all blocks are combined into the compressed matrix $DCT^{comp} = (DCT^{comp}_{k'l'})$ of sizes $K' \times L'$ $(l' = 0,1,..., L'-1,\ k = 0,1,..., K-1')$;
- an integer from the interval [0,255] is assigned to each element of DCT^{comp} by the formula as follows:

$$w'_{k'l'} = \left[255 \frac{DCT^{comp}_{k'l'} - Min}{Max - Min} \right], \qquad (1)$$

where Max, Min are the maximal and minimal elements of the matrix DCT^{comp}.

Thus, the matrix (1) can be considered as the matrix W' of pixels intensities of the spectral pattern of embedding watermark (SPW).

2.1.2. Embedding the spectral pattern of the watermark

- the original image to protect $I = \{a_{mn}\}$ of sizes $M \times N$, $m = 0,1,... M-1$, $n = 0,1,... N-1$ is splitting into $K' \times L'$ nonoverlapped blocks I_ξ by the number of pixels of the SPW $w'_{k'l'}$ of sizes $(M/K') \times (N/L')$, where $\xi = 0,1,2..., (K' \times L') - 1$;
- if M and N are not multiples of K' and K' respectively, SPW is complemented with necessary number of zeros;
- each block of the image I is placed in one-to-one correspondence with pixels of the transformed watermark W'. Depending upon the particular application may be applied algorithm of mixing the coordinates of pixels with a certain rule, for example, with Arnold transform;
- let w'_ξ be ξ-th pixel of watermark W'. Let pixel w'_ξ be embedded into block $I_\xi = \{a_{m_\xi n_\xi}\}$. The embedding procedure proceeds by formula as follows:

$$a^{W'}_{m_\xi n_\xi} = (1-\alpha) a_{m_\xi n_\xi} + \alpha w'_\xi,$$

where $a^{W'}_{m_\xi n_\xi}$ is corresponding pixel of watermarked image $I^{W'}$ at fixed $\alpha > 0$ for all blocks.

2.2. Watermark extracting algorithm

Denote the attacked watermarked image by $I^{W'X} = \{a^{W'X}_{ij}\}$, where X denotes the attack. The watermark extraction algorithm proceeds as follows:

- image $I^{W'X} = \{a^{W'X}_{ij}\}$ is splitted into blocks by number of pixels of embedded watermark. Denote the average intensities of pixels of blocks of the image-container by $\{\mu_\xi\}$, where

$$\mu_\xi = \frac{KL}{MN} \sum_{m_\xi} \sum_{n_\xi} a_{m_\xi n_\xi};$$

- denote the average intensity of pixels of ξ-th block of the attacked watermarked image-container by $\mu^{W',X}_\xi$, where

$$\mu^{W',X}_\xi = \frac{KL}{MN} \sum_{m_\xi} \sum_{n_\xi} a^{W',X}_{m_\xi n_\xi};$$

- the estimation of an element of SPW at presence of attack X is calculated by formula as follows:

$$\hat{w}^{,X}_\xi = \frac{\mu^{W',X}_\xi - (1-\alpha)\mu_\xi}{\alpha} \qquad (3)$$

Formula (3) is obtained by the least-squares method.

- the extracted data are transformed by formula as follows:

$$DCT^{comp}_{k'l'} = w'_{k'l'} \frac{Max - Min}{255} + Min; \qquad (4)$$

- the matrix with elements (4) is splitted into blocks of sizes $t \times t$, and each block is complemented with necessary number of zeros up to sizes of $b \times b$;
- each block undergoes the inverse DCT;
- the data are combined in the matrix and converted into an image with sizes of initial size of watermark;
- the pixels of the image are mixed according to the inverse rule, which was applied at embedding process.

The combined spatial-frequency watermarking algorithm described above allows to manipulate images with different aspect ratio of the image container and watermark, which is especially important in problems of the detection of forgeries with unknown in advance geometry.

3. EXPERIMENTAL RESULTS

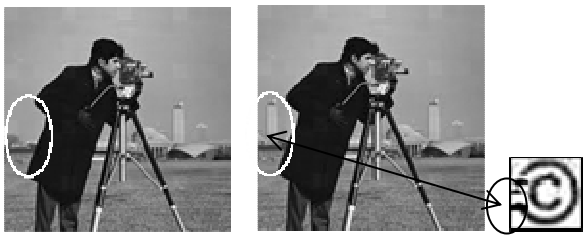
To test the effectiveness of the combined algorithm described above a series of numerical experiments is carried out which used as a protection object test image Cameraman of size 256 × 256 pixels (Fig. 1a), as a watermark image a logo of size 32 × 32 pixels (Fig. 1b) and a fragment of image-container of sizes 128 × 128 pixels (Fig. 1c).

a b c

Figure 1. Images which are used in the experiments: image-container (a), watermark-logo (b), a fragment of image-container, used as the watermark (c).

3.1 Experiment 1

To control the integrity of the original image Cameraman (Fig. 1a) it was watermarked by the watermark-logo. Since the purpose of embedding the watermark is not only the finding of fraud, but the determination of the fragment which has been tampered, the algorithm of mixing of pixels in this experiment does not apply. Resulting eventually protected image (Fig. 2a) has similarity to the original PSNR = 37.34 dB.



a b c

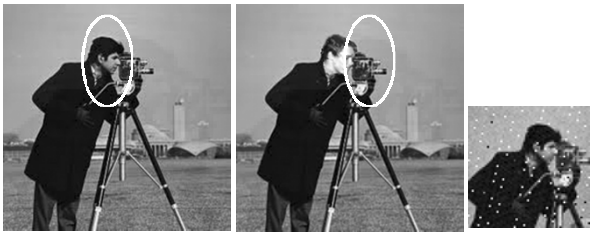
Figure 2. Results of Experiment 1: Protected image (a), tampered image (b), extracted watermark (c).

Then a fragment of protected image is copied to the same image (Copy-Paste attack). In Figure 2b the obtained image is shown. After watermark extracting by combined algorithm we have the image of Fig. 2c, and the location of its distorted part indicates the distorted fragment of the protected image.

Certain regions of an image can present special interest from the point of view of investigation or protection. Such regions are called Region of Interest (ROI). Therefore it is very important to have a possibility to recover the distorted ROI if the attack takes place. To provide such possibility the ROI-image can be embedded into the container as a watermark using the combined watermarking procedure.

3.2. Experiment 2

A fragment of image Cameraman is embedded to itself at $\alpha = 0.03$ after preliminary compressing up to sizes of 64x64 pixels (Fig. 1c). As a result we get the protected image of Fig. 3a with PSNR=39.05 dB. Then a modelled attack of type "Collage" is applied, so we get the tampered image of Fig. 3b. One can see that the extracted watermark though is distorted but allows getting some impressions about the content of the original image.



a b c

Figure 3. Results of Experiment 2. Protected image (a), Tampered image (b), Extracted watermark (c).

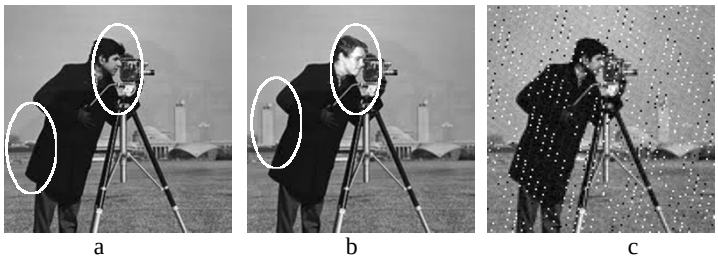
3.3. Experiment 3

The images with multiple ROI can be found in practice. To protect and subsequently recovery such images the embedding a copy of the whole image into the original one is proposed. It can be performed by using the combined algorithm.

In this experiment a copy of this image is self-embedded into the image Cameraman at $\alpha = 0.03$. The protected image with PSNR= 37.53 dB is shown in Fig. 4a.

After application of modelled attack of type "Collage" we get an image with two tampered fragments (Fig. 4b). One can see that the extracted watermark though is distorted but allows to recover partially the lost information about the original image.

Thus, embedding a watermark using the combined technique allows establishing the fact of tampering of protecting image and determining the distorted fragment. It is possible to recover partially the content of the initial image by comparing the extracted watermark and the distorted image.



a b c

Figure 4. Results of Experiment 3. Protected image (a), Tampered image (b), Extracted watermark (c).

4. CONCLUSIONS

In this paper, a problem of unauthorized changing of the content of an image, determination of the location of distorted parts and partial recovery of distorted parts is considered. The most commonly used types of fraud are based on operations such as Copy-Paste, Copy-Move and its combinations. The common technique for detection of such attacks, used in many investigations, is based on using a watermarking procedure in such a way that the type and parameters of distortion of the watermark from an attack allow to make an inference on presence of a fraud, to locate the damaged parts of the image and even to recover them. In this paper, the investigated earlier combined spatial-frequency watermarking algorithm is used, which allows the manipulations over the images of different ratios of image-container and watermark sizes. Possibility of self-embedding of an image container or its fragment by using the proposed combined algorithm is shown. Different schemes of image tampering and partial recovery of distorted parts of an image are suggested. Numerical experiments show the effectiveness of the proposed schemes and algorithms.

REFERENCES

[1] I.J. Cox, M.L. Miller, and J.A. Bloom. Digital Watermarking. Morgan Kaufmann Publishers, 2002.

[2] Chen-Kuei Yang, Chang-Sheng Huang. “A Novel Watermarking Technique for Tampering Detection in Digital Images”. *Electronic Letters on Computer Vision and Image Analysis*, 3(1): 1-12, 2004.

[3] S. Dadkhah, A. A. Manaf, S. Sadeghi. “Efficient Digital Image Authentication and Tamper Localization Technique Using 3LSB Watermarking”. *International Journal of Computer Science*, Vol. 9, Issue 1, No 2, January 2012.

[4] D.G. Asatryan, N.S. Lanina, H.S. Shahverdyan. “Adaptive Robust Algorithm for Digital Watermarking of Medical Images”. *Proc. of 6-th Int. Conf. on Computer Science and Information Technologies – CSIT'2007*, Yerevan, pp. 161-164, 2007.

[5] J. Fridrich and M. Goljan. “Images with self-correcting capabilities”. In *IEEE International Conference on Image Processing*, vol 3. Kobe, Japan, pp 792–796 1999.

[6] X. Zhu, A.T.S. Ho, and P. Marziliano. "A New Semi-fragile Image Watermarking With Robust Tampering Restoration Using Irregular Sampling," Elsevier Signal Processing: Image Communication, Vol. 22, Issue 5, p515-528, June 2007.

[7] Chun-Wei Yang, Jau-Ji Shen. “Recover the tampered image based on VQ indexing”. *Signal Processing*, Vol. 90, Issue 1, pp. 331-343, 2010.

[8] D. G. Asatryan, N. S. Asatryan. “Combined Spatial and Frequency Domain Watermarking”. *Proc. of 7-th Int. Conf. on Computer Science and Information Technologies – CSIT'2009*, Yerevan, pp. 323–326, 2009.

[9] H. Rahmani, R. Mortezaei, M.E. Moghaddam. “A New Robust Watermarking Scheme to Increase Image Security”. *EURASIP Journal on Advances in Signal Processing*. Vol. 2010.