A CONVOLUTIVE MIXING MODEL FOR SHIFTED DOUBLE JPEG COMPRESSION WITH APPLICATION TO PASSIVE IMAGE AUTHENTICATION

*Zhenhua Qu, Weiqi Luo, Jiwu Huang**

Guangdong Key Lab. of Information Security Technology
Sun Yat-Sen University, Guangzhou, China, 510275

ABSTRACT

The artifacts by JPEG recompression have been demonstrated to be useful in passive image authentication. In this paper, we focus on the shifted double JPEG problem, aiming at identifying if a given JPEG image has ever been compressed twice with inconsistent block segmentation. We formulated the shifted double JPEG compression (SD-JPEG) as a noisy convolutive mixing model mostly studied in blind source separation (BSS). In noise free condition, the model can be solved by directly applying the independent component analysis (ICA) method with minor constraint to the contents of natural images. In order to achieve robust identification in noisy condition, the asymmetry of the independent value map (IVM) is exploited to obtain a normalized criteria of the independency. We generate a total of 13 features to fully represent the asymmetric characteristic of the independent value map and then feed to a support vector machine (SVM) classifier. Experiment results on a set of 1000 images, with various parameter settings, demonstrated the effectiveness of our method.

Index Terms— Image Forgery, Shifted Double JPEG, Convolutive Mixing Model, ICA, Image Authentication

1. INTRODUCTION

With the rapid advancement in digital image editing software, the verification the trustworthy of a given digital image is receiving increasing concerns, especially in the legal system and journalism. The traditional approach of image authentication is to use digital signature and digital watermarking. However, these two methods will fail if the signature or watermark is not employed prior to the malicious editing of the untampered image. Recently, some researchers tried to address a new approach, called the *Passive Image Authentication* which only uses the characteristic from the forgery/fake images to figure them out. Some pioneer works are done by Farid et al. and Fridrich et al. [1][2].

The JPEG recompression artifacts is one kind of those important characteristics. It is unlikely for an ordinary JPEG image/photo or a part of it to possess recompression artifacts, unless it has been opened and re-saved as JPEG format by an image editing software. Recent researches have been focusing on the *Double JPEG Compression Problem*[1], which means to identify the image suffered from the lossy JPEG compression twice. A possible solution for this problem presented by Lukas and Fridrich [2] is to estimate the primitive quantization table coefficients from a double compressed JPEG image directly. Other approaches tried to detect the abnormal DCT coefficient distribution caused by the recompression. In [1], Popescu and Farid utilized the periodical artifacts of the re-quantized DCT coefficient histogram. Fu et al. [3] contributed another identification algorithm based on a generalized Benford Law model of the blockwise discrete cosine transform (BDCT) coefficients. However the above methods did not address how to determine whether a given JPEG image had been recompressed with inconsistent block segmentation which frequently occurs in a composite or region-duplication image. We define the task of identifying such a JPEG recompressed image with *Shifted Double JPEG (SD-JPEG) Compression Problem*. He et al. [4] noticed the difference between these two types of problem and utilized it to detect forgery, but they did not derive any explicit model for the SD-JPEG problem.

In previous work [5], we proposed an rapid identification algorithm in spatial domain based on the symmetric characteristic of blocking artifact characteristics matrix (BACM). In this paper, our contribution comes in three aspects : 1) We address a statistic model in the BDCT domain for the SD-JPEG problem and give it a new insight from the direction of BSS. 2) An novel ICA-based identification algorithm utilizing the symmetric property of IVM is proposed. 3) The problem of estimating the shifted distance is also addressed here to provide detailed information of how the image is recompressed.

The rest of this paper is organized as follows. In Section 2, we establish a convolutive mxing model of SD-JPEG compression. Section 3 mainly discusses the identification of the model with a ICA-based method and how to estimate the shifted distance. The experiment results with comparison to

*This work is supported by NSFC (90604008, 60633030), NSF of Guangdong (04205407), and 973 Program (2006CB303104).
Contact Author (isshjw@mail.sysu.edu.cn)

[1]Other than this literal explanation, this term has a more specific meaning which stands for unshifted double JPEG compression problem in this paper.

the former solution are shown in Section 4 and future works will be presented in Section 5.

2. BACKGROUND

In this section we firstly give a brief introduction of image-splicing model in image forgery which serves as a background of the SD-JPEG problem. Then we will establish a statistical model for SD-JPEG compression.

2.1. Splicing Model for JPEG Image

Image-splicing is a most common manipulation technique in image forgery. It means that a part of the source image is copied to the target image to generate a new composite image, as illustrated in Figure 1.

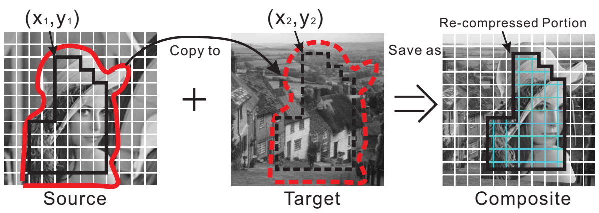

Fig. 1. JPEG image splicing model.

Supposing that both the source and composite images are stored in JPEG format, with quantization factor QF_1 and QF_2 respectly. The image block in the spliced portion constituted by a number of JPEG blocks which originally located at (x_1, y_1) in the source image is copied and pasted to (x_2, y_2) in the target and composite images. The *Shifted Distance* (cx, cy) of an image block is defined as $cx \triangleq (x_1 - x_2) \mod 8$, $cy \triangleq (y_1 - y_2) \mod 8$.

Since the composite image in suspicion can be segmented into small blocks and identified one by one, in this paper we will only address the problem of identifying a SD-JPEG compressed image from an ordinary JPEG compressed one.

2.2. The Convolutive Mixing Model of SD-JPEG

Shown in Fig. 2, the SD-JPEG compressing process generates an SD-JPEG image in three steps: 1) Decompress an ordinary JPEG (only compressed once). 2) Crop off cx columns and cy rows in spatial domain. 3) Recompress it.

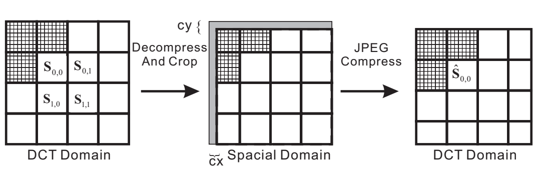

Fig. 2. Generation of an SD-JPEG image.

The input and output of this process are $M \times M$ DCT coefficient blocks denoted with $\mathbf{S}_{m,n}$ and $\hat{\mathbf{S}}_{m,n}$ respectly. The linear characteristic of DCT transform indicates that an output block $\hat{\mathbf{S}}_{m,n}$ can be expressed as a linear mixture of four input blocks $\{\mathbf{S}_{m,n}, \mathbf{S}_{m,n+1}, \mathbf{S}_{m+1,n}, \mathbf{S}_{m+1,n+1}\}$ that it overlaps. The input-output relationship of SD-JPEG compression is

$$\widehat{\mathbf{S}}_{m,n} = \sum_{i=0}^{1}\sum_{j=0}^{1} \mathbf{A}_{cy,i}\mathbf{S}_{m-i,n-j}\mathbf{A}_{cx,j}^{T} + \hat{\mathbf{E}}_{m,n} \qquad (1)$$

where $\hat{\mathbf{E}}_{m,n}$ is the quantization noise of the second JPEG compression. $\{\mathbf{A}_{cx,0}, \mathbf{A}_{cx,1}, \mathbf{A}_{cy,0}, \mathbf{A}_{cy,1}\}$ are called a set of mixing matrices. Their coefficients are determined by the shifted distance (cx, cy) and the DCT transform matrix. We will not go into the detailed derivations of them, because the multiplications of these matrices are done implicitly in the aforementioned generation process by a JPEG compression software. Equ. (1) is a 2D convolutive mixing model mostly studied in BSS.

2.3. An ICA-based Solution

To identify if a given JPEG image is SD-JPEG compressed, we need to blindly estimate the de-mixing matrices and thus can determine their corresponding shifted distance. *If the estimated shifted distance is not (0,0), the image is an SD-JPEG compressed one*. Intuitively, the ICA algorithm is suitable for this task. It works by iteratively optimizing some de-mixing matrices as to separate a number of mixed signals (input components) into independent output components. The independency between the output components is measured by an objective function [6].

There are some pros and cons should be noted in practice. The main difficulty is how to accurately estimate the ground truth shifted distance using the ICA algorithm, especially when the additive noise $\hat{\mathbf{E}}_{m,n}$ is very strong. We have to derive an alternative rule based on above rule to make judgment. This may cause a few undetectable conditions. But we also have an unique advantage over conventional ICA algorithms. We notice that the de-mixing of Equ. (1) is achieved when the SD-JPEG image is shifted back to its original block segmentation. For example, if an SD-JPEG compressed image is generated by shifting (cx, cy), then a further shift of $(M - cx, M - cy)$ can de-mixing it. Thus the number of all possible de-mixing matrix set is limited and identical to the number of all possible shifted distances. When the block size is $M \times M = 8 \times 8$, this number is only 64. Owning to this advantage, the iteratively optimized ICA algorithm can be reduced to an exhaustive search among a total of 64 objective function values.

3. PROPOSED METHOD

We propose an ICA-based method which is composed of three steps: First the objective function values of all possible de-mixing matrix set are calculated to form an IVM. Then the relative asymmetric value map (RAVM) is derived to measure the symmetry of the IVM. Finally a judgment is made by a classifier trained with features extracted from the ordinary/SD-JPEG images.

3.1. Calculation of the Independent Value Map

Supposing that an SD-JPEG compressed image is de-mixed by shifting (x, y), we treat the resulted DCT coefficients within the $M \times M$ blocks as random variables and denote them as $\left\{\mathbf{s}_{i,j}^{x,y} \middle| i, j = 0...M-1\right\}$. The objective function used to measure the independency between these output components is proposed by Hyvärinen [6].

$$I(\mathbf{s}_{0,0}^{x,y}, \mathbf{s}_{0,1}^{x,y}, ..., \mathbf{s}_{M-1,M-1}^{x,y}) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} J(\mathbf{s}_{i,j}^{x,y}) \quad (2)$$

We choose kurtosis as $J(\mathbf{s})$, thus $J(\mathbf{s}) = \left[\frac{E(s-\mu)^4}{\sigma^4} - 3\right]^2$ where μ and σ are the mean and standard deviation of random variable $\mathbf{s}$. Consequently, the IVM is calculated as follows.
Input: A JPEG compressed image in suspicion
Output: Independent Value Map $\mathbf{I}$
i) Decompress the JPEG image into spatial domain.
ii) Perform BDCT on the decompressed image with 64 different segmentation schemes. For each scheme denoted by $(x, y), x, y = 0, 1, ..., M-1$, we obtain an independent value of the output components $\left\{\mathbf{s}_{i,j}^{x,y} | i, j = 0, ..., M-1\right\}$ by evaluating Equ. 2 and store it into the $M \times M$ *Independent Value Map* $\mathbf{I}$.

$$\mathbf{I}(x, y) = I(\mathbf{s}_{0,0}^{x,y}, \mathbf{s}_{0,1}^{x,y}, ..., \mathbf{s}_{M-1,M-1}^{x,y})$$

In practice, the terms $J(\mathbf{s}_{i,j}^{x,y})$, $i, j < 2$, may be omitted for their little contribution to the identifiability.

As shown in Fig. 3, large independent values appear around de-mixing shifted distance $(M-cx, M-cy)$. The notable difference between ordinary JPEG and SD-JPEG image is their *symmetry*. Note that all IVMs shown below are circularly shifted for (3,3), e.g. the value $\mathbf{I}(0, 0)$ is shown at (3,3) now.

3.2. Symmetry of IVM

Referring to Fig. 4, the IVM is divided into 3 regions $\Re_1$,$\Re_2$ and $\Re_3$, labeled with red, green and blue. For ordinary JPEG images, the independent values within each region should be centrosymmetric according to their center block labeled with red border. *If this symmetry is disrupted , the image is highly suspected to be an SD-JPEG image.*

The three (except the one in $\Re_1$ representing the unshifted double JPEG compression $(cx = 0, cy = 0)$) crossed over blocks are currently undetectable in this work, since they have no centrosymmetric counter parts to compare with.

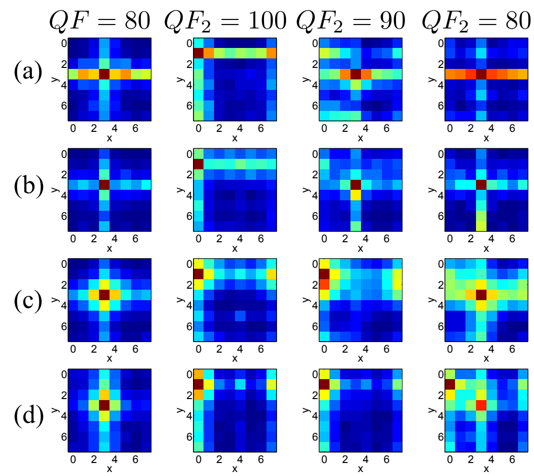

Fig. 3. IVMs of four standard images: (a)lena (b)mandrill (c)barb (d)goldhill. The first column are the IVMs of ordinary JPEG images with $QF = 80$. Others are of SD-JPEG images with $QF_1 = 70$, $QF_2 = 100, 90, 80$ and the shifted distance is $(cx = 3, cy = 2)$. All IVMs are plotted in log-scale with warmer color for larger independent value.

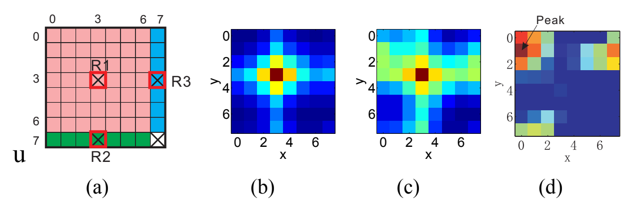

Fig. 4. Symmetry of IVM. (a)Three regions and their centers. (b)Symmetric IVM of an ordinary JPEG image. (c)Asymmetric IVM of an SD-JPEG image. The IVMs in (b) and (c) are chosen from the barb image in Fig. 3. (d)RAVM calculated from (c).

We define the Relative Asymmetric Value Map as a normalized descriptor of the of IVM.

$$\mathbf{R}(i, j) = \frac{\max\{\mathbf{I}(i, j) - \mathbf{I}[f_c(i, j)], 0\}}{\max\{\mathbf{I}(i, j), \mathbf{I}[f_c(i, j)]\}} \quad (3)$$

where $\mathbf{I}[f_c(i, j)]$ is the centrosymmetric counter part of $\mathbf{I}(i, j)$ according to the center block of the region it belongs to. By inspecting Fig. 4(d), the maximum relative asymmetric value (RAV) is located at $(x_{max} = 5, y_{max} = 6)$ (before circular shift), then ground truth shifted distance $(3, 2)$ is obtained by

$$(cx, cy) = (8 - x_{\max}, 8 - y_{\max}) \quad (4)$$

3.3. Automatic Identification Applying SVM Classifier

To make the identification process automatically, we adopt a two-class SVM classifier to distinguish the SD-JPEG compressed image form ordinary one. As to fully exploit the discriminative power of RAVM pattern, 13 features are extracted : Maximum RAV of $\mathbf{R}$ (1 feature) measures the peak in RAVM caused by SD-JPEG compression as shown in Fig. 4(d). Centroid of Each Region (4 features, two for $\Re_1$, one for $\Re_2$, one for $\Re_3$) and Circular Centroid of RAVM (4 features, two for horizontal, two for vertical) measure the shift of centroid caused by SD-JPEG compression. Sum of RAV in each region (3 features) captures that in strong noise the maximum of RAV may not be significant but the sum of RAV is still large. Maximum Number of Circular Connected Blocks (1 feature) depicts that large RAVs (> 0.1 in experiment) tend to cluster together for the RAVM of SD-JPEG image. And the connectivity should be considered in a circular way.

4. EXPERIMENTAL RESULTS

Our image database [5] includes 1000 images taken by a Panasonic DMZ-FZ30 digital camera with a variety of indoor and outdoor scenes. The original image resolution is 3264×2448 stored in uncompressed TIFF format. We test performance of the algorithm under a combination of three image sizes (640×480, 1280×1024 and 1600×1200 obtained by cropping the TIFF image and saved as grayscale bmp) and eight noise levels ($QF_2 \in [60, 65, ..., 95]$).

In a specified test condition, we generate a pair of ordinary JPEG image (authentic) and SD-JPEG image (tampered) for each bmp image with strict consistency both in size and image contents. The SD-JPEG images are generated as described in Sec. 2.1 and illustrated in Fig. 2. For ordinary JPEG images, we crop the same portion of the bmp image and simply save them as JPEG format with QF_2.

As a result, for every test condition we obtain 2000 images. We randomly pick out half of them (500 ordinary JPEG images and 500 SD-JPEG images) for training and others for testing. Then the 13 features extracted from the RAVM are fed to an SVM classifier with radial basis kernel implemented in LIBSVM [7]. This process is repeated for five times to obtain the average performance shown in Fig. 5 with comparison to the BACM-based algorithm in [5] .

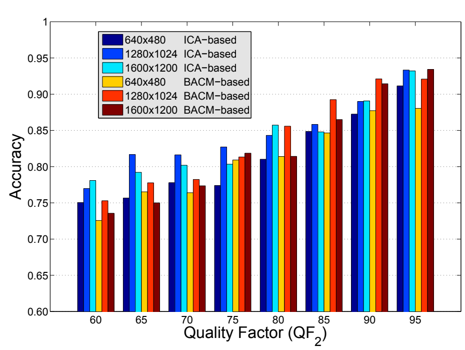

Fig. 5. Comparison of ICA-based and BACM-based methods.

Overall, the performance of two different algorithms are quite close. In high noise level ($QF_2 \leq 70$) , the ICA-based algorithm is more robust than the BACM-based one. Our results also improve the accuracy in the worst case ($QF_2 = 60, size = 640 \times 480$) to about 75%. The main insight here is that our BDCT domain convolutive mixing model performs at the same level as existing spatial domain model and a further improvement of existing method is possible.

5. CONCLUSION AND FUTURE WORK

To deal with the challenging image forgery detection, we contribute a convolutive mixing model for better interpreting the SD-JPEG problem. We find that the presence of SD-JPEG compression would weaken the independency between BDCT coefficients. This also leads to a novel ICA-based identification algorithm which can be applied to detect JPEG image-splicing forgery. As for perspectives, the model and method developed in this work may also be extended to color JPEG images and other blockwise compressed multimedia, such as JPEG2000. These are our future works.

6. REFERENCES

[1] A.C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Proc. of the 6th International Workshop on Information Hiding*, Toronto, Canada, May 2004, vol. 3200 of *LNCS*, pp. 128–147.

[2] J. Lukas and J. Fridrich, "Estimation of primary quantization matrix in double compressed jpeg images," in *Proc. Digital Forensic Research Workshop*, Cleveland, OH, USA, Aug. 2003.

[3] D.-D. Fu, Y.-Q. Shi, and W. Su, "A generalized benford's law for jpeg coefficients and its applications in image forensics," in *Proc. SPIE Security, Steganography, and Watermarking of Multimedia Contents*, 2007, vol. 6505, p. 58.

[4] J. He, Z. Lin, L. Wang, and X. Tang, "Detecting doctored jpeg images via dct coefficient analysis," in *ECCV'06*, 2006, vol. III: 423-435.

[5] W.-Q. Luo, Z.-H. Qu, J.-Wu. Huang, and G.-P. Qiu, "A novel method for detecting cropped and recompressed image block," in *Proc. IEEE ICASSP'07*, Hawaii, 2007.

[6] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *IEEE Trans. Neural Networks*, vol. 13, no. 4-5, pp. 411–430, Jun. 2000.

[7] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM: a library for support vector machines*, 2001, Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm.