# Markov-based Image Forensics for Photographic Copying from Printed Picture*

Jing Yin[1] ,Yanmei Fang [2]
School of Information Science & Technology, Sun Yat-sen University
Guangzhou, 510006, China
[2]fangym@mail.sysu.edu.cn

## ABSTRACT

Nowadays, photographic-copying technique is very popular along with the rapid development of the image-capturing device, especially digital camera. As a result, the recaptured images, i.e., images taken from real-scene images displayed on various medium, e.g., LCD screen, are used in illegal cases now and then.In this paper, by comparing the recaptured images with their corresponding real-scene images, we find the recapturing procedure changes the statistics of the images. Then the Markov process based features extracted from the Discrete Cosine Transform(DCT) coefficients array are proposed to characterize this changes. During experimentation, a large and typical image dataset, which consisted of 3994 real-scene images and 3994 recaptured images that are taken from printed pictures with diversified image contents and camera models, is build and used for training and testing the classifier Support Vector Machine(SVM). Experimental results show that the proposed forensics scheme performs very well and outperforms the state-of-art methods.

## Categories and Subject Descriptors

I.4 [**Image Processing**]: Miscellaneous

## General Terms

Security, Algorithms

## Keywords

Digital image forensics, photographic copying, Markov process, SVM

## 1. INTRODUCTION

Image-displaying technology has greatly developed over the past few decades, thus we can see high-quality images displayed on LCD screen, poster and printing paper everywhere. If we use the image capturing device to recapture those displayed images, we will get the copies of those images. This process, called photographic copying, can be modeled as Fig.1. Photographic copying is a special case of creating images, i.e., both the input and output of image capturing device are images. The recaptured images have the same content with the corresponding real-scene images, but the image capturing device are different, and the information recorded in the header of the image are "renewed". What's worse, the displayed images may be tampered. As the results of photographic copying, the tampering traces which can be used to identify tampering by current forensics methods were automatically wiped. There was a real example in our daily life [1]. In 2007, a Chinese farmer claimed to have taken 71 photographs of the almost extinct South China tiger, and then this news aroused a large-scale controversy about the authenticity of those images. However, after the government's investigation, this farmer confessed his guilt that his photos of tiger were recaptured from a life-size-poster. Another threat brought by recaptured images is rebroadcast attack. Nowadays, face recognition technologies are widely used on consumer electronics. However, an article from CNET [2] demonstrated that current face recognition systems deployed in the market for consumers are extremely vulnerable to playback attacks. What's important, identifying images obtained by photographic copying make senses in the area of image source forensics.

Some researches have been presented for identifying recaptured images in the literature. The research of Yu et al.[3] and Gao et al.[4] both focus on identifying recaptured images obtained by photographing the print-out of the image on a paper. Yu et al.[3] found that the recaptured images and their original images are different in the distribution of the gradient of specular ratio image. This find was also explored in [5] to solve the playback attacks in face recognition system. Gao et al.[4] used a set of physics-based features, which exploited the contextual information on remaining on a recaptured image and the unique properties of the recaptured image rendering process, to differentiate the recaptured images from the real-scene images. Cao et al.[6] studied the detection of finely recaptured photos on common LCD screen. The features used in their research included texture features, loss-of-detail features and color features. In [7], two sets of features based on the noise and the traces of double JPEG compression are proposed to identify recaptured images taken from LCD screen.

In this paper, the photographic copying considered is that

*This work was supported in part by 973 Program of China (2011CB302204) and NSFC (U1135001)


*MM'12,* October 29–November 2, 2012, Nara, Japan.

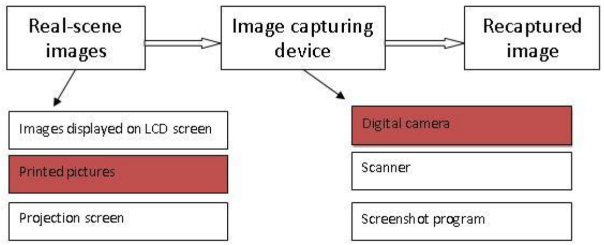


**Figure 1: What is photographic copying.**

digital camera recapture printed pictures. Markov process based features are proposed to classify real-scene images and recaptured images taken from printed pictures. The experiments conducted on a generalized image database show that the proposed features perform very well for detecting recaptured images taken from printed pictures and outperform the state-art methods. The rest of this paper is organized as follows. In Section 2, the proposed markov based features was described. In Section 3, the experimental results and discussions are presented. In Section 4, we draw a conclusion for this paper and make a plan for the future research.

# 2. THE PROPOSED FEATURES

In photographic copying, an image is firstly printed on glossy inkjet photo paper, thus the dithering that results from printing disturbs the natural statistic of images to some extent. By comparing recaptured images with their corresponding real-scene images, we find the differences between them. Firstly, the details of recaptured images are more blurry, which indicates that the correlations among the pixels in recaptured images are more weak. Secondly, the color of recaptured images and their corresponding real-scene images are a little different. This difference may relate to the technique of printing and the environment of photographic copying. Motivated by the above observation, we expect that some natural image statistic models can be used to separate recaptured images from real-scene images. The natural image statistic features used in this paper are Markov transition probabilities extracted from the DCT coefficients array of the test image. The Markov based features are firstly proposed in [8] as an approach to effective attacking JPEG steganography. The details of feature generation will be described in the following section.

## 2.1 The DCT Coefficients

The format of the images in our database is JPEG, which is widely used by the memory system of digital cameras. For a JPEG image, the JPEG coefficients(i.e., the quantized DCT coefficients) are considered. The procedure of photographic copying changes the statistics of images, i.e., the correlation among elements. The JPEG coefficients can reflect this change. A JPEG color image has three components, i.e.,$Y, C_b$ and $C_r$. We only consider the coefficients in Y component for the following two reasons. Firstly, as $C_b$ and $C_r$ have been downsampled during JPEG compression, some information have been lost. Secondly, if all the components are considered, computational complexity will be extremely high.

As mentioned earlier, we find that the color of recaptured images are different with the real-scene images. Besides the $YC_bC_r$ color model, RGB color model is also widely used to represent color images. Actually, when an image is JPEG compressed inside the digital camera, the image is firstly converted to $YC_bC_r$ color model from RGB color model. So we also consider the DCT coefficients in RGB color model. In order to reduce computational complexity, we only consider the R component. We get the DCT coefficients of R component by the following three simple steps:(1)Convert the JPEG color image to RGB color model; (2)Apply 8×8 block discrete cosine transform to the R component to obtain the DCT coefficients; (3)Round each coefficient to the nearest integer to obtain the DCT coefficients array.

In summary, the DCT coefficients considered in this paper include the quantized DCT coefficients of the image in Y component and the DCT coefficients array obtained by applying block discrete cosine transform to the R component of the image(i.e.,the DCT coefficients in R component). Further discussion about which kind of DCT coefficients array is more useful will be made in Section 3.

## 2.2 Feature Generation

### 2.2.1 *Difference DCT Coefficients Array*

In order to reduce the effect caused by the diversity of images content, we introduce the difference DCT coefficients array, i.e., the difference between an elements and its neighbor elements in an DCT coefficients array. We denote the absolute value of the DCT coefficients arrays by $F(u, v)(u \in [0, D_u-1], v \in [0, D_v-1])$, where $D_u \times D_v$ is the size of image. The difference DCT coefficients arrays, i.e., $F_h(u, v)$,$F_v(u, v)$, $F_d(u, v)$, $F_{md}(u, v)$, are computed as the following formulas:

$$F_h(u, v) = F(u, v) - F(u+1, v) \tag{1}$$

$$F_v(u, v) = F(u, v) - F(u, v+1) \tag{2}$$

$$F_d(u, v) = F(u, v) - F(u+1, v+1) \tag{3}$$

$$F_{md}(u, v) = F(u+1, v) - F(u, v+1) \tag{4}$$

where $u \in [0, D_u-2], v \in [0, D_v-2]$) and $F_h(u, v)$,$F_v(u, v)$, $F_d(u, v)$,$F_{md}(u, v)$ denote difference DCT coefficients array in the horizontal, vertical, main diagonal and minor diagonal direction, respectively.

### 2.2.2 *Transition Probability Matrix*

In order to capture the changes resulting from the photographic copying, the difference DCT coefficients array is modeled as a Markov random process and the one-step transition probability matrix (TPM) is computed to characterize this Markov process[9]. With the purpose of reducing computational complexity, we resort to a thresholding technique as given in the following formula (5):

$$F_h^T(u, v) = \begin{cases} T, & \text{if } F_h(u, v) > T. \\ F_h(u, v), & \text{if } \|F_h(u, v)\| \le T. \\ -T, & \text{if } F_h(u, v) < -T. \end{cases} \tag{5}$$

where $T$ is the threshold value. The above formula means that the element which is lager than $T$ will be represented by $T$, and the element which is smaller than $-T$ will be represented by $-T$. The same thresholding technique is also conducted on the other three difference DCT coefficients array. Therefore, we obtain four new difference DCT coefficients arrays, i.e., $F_h^T(u, v), F_v^T(u, v)$,$F_d^T(u, v)$ and $F_{md}^T(u, v)$. The one-step transition probability matrix associated with the horizontal difference DCT coefficients array is given in Eq.(6):

$$P^h = \begin{array}{c} \\ \text{-T} \\ \text{-T+1} \\ \vdots \\ \text{T} \end{array} \begin{array}{c} \begin{array}{cccc} \text{-T} & \text{-T+1} & \dots & \text{T} \end{array} \\ \begin{pmatrix} p_{11}^h & p_{12}^h & \dots & p_{1n}^h \\ p_{21}^h & p_{22}^h & \dots & p_{2n}^h \\ \vdots & \vdots & & \vdots \\ p_{m1}^h & p_{m2}^h & \dots & p_{mn}^h \end{pmatrix} \end{array} \quad (6)$$

where $m = n = 2T + 1$, and the elements of this TPM are computed as Eq.(7):

$$p_{ij}^h = p(F_h^T(u+1,v) = j - T - 1 \mid F_h^T(u,v) = i - T - 1) = \frac{\sum_{v=0}^{D_v-2}\sum_{u=0}^{D_u-2} \delta(F_h^T(u,v) = i - T - 1, F_h^T(u+1,v) = j - T - 1)}{\sum_{v=0}^{D_v-2}\sum_{u=0}^{D_u-2} \delta(F_h^T(u,v) = i - T - 1)} \quad (7)$$

where $i \in \{1, 2, \cdots, 2T+1\}$, $j \in \{1, 2, \cdots, 2T+1\}$, and

$$\delta(A) = \begin{cases} 1, & \text{if } A \text{ is true.} \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

The one-step TPMs associated with vertical, main diagonal and minor diagonal difference DCT coefficients array can be computed in the similar way. Thus we obtain $(2T + 1) \times (2T + 1)$ elements for each of these four TPMs. All of these elements are used as our proposed features. The discussion on the selection of threshold $T$ will be presented in the section 3.

# 3. EXPERIMENTS AND RESULTS

The classifier used in our experiments is support vector machine (SVM) with RBF (radial basis function) kernel. The Matlab codes of SVM can be downloaded from [10]. For each experiment, 5/6 of the recaptured images (positive set) and 5/6 of real-scene images (negative set) are randomly chosen to train the SVM classifier, and the remaining 1/6 of recaptured images and 1/6 of real-scene images are used for testing the trained classifier. The classification results, which are averaged over 20 times for random cross-validation, are presented in each experiments. In each table, TP (true positive) represents the detection rate of recaptured images, TN (true negative) represents the detection rate of real-scene images, and the accuracy is the arithmetic average of TP and TN.

## 3.1 Experiments on the Generalized Database

In order to test the effectiveness of the proposed features, we set up an generalized database comprised of 3994 recaptured images taken from printed pictures and 3994 real-scene images. The printed pictures used in the photographic copying include 348 pictures printed on 4R glossy photo papers and 261 pictures on magazines. These two kinds of pictures are most common in our daily life. 7 different digital cameras are used in the photographic copying and some poor recaptured images are rejected. The real-scene images are also taken by 7 different cameras. Among these 3994 real-scene images, 348 images are the original images of the printed pictures and the rest are collected by our team members. All of the images in this generalized database are saved in JPEG format and their resolutions are greater than $2200 \times 1700$.

When we extract the Markov-based features from the DCT coefficients in Y components, the experimental results with different threshold T are shown in Table 1. When we extract the Markov-based features from the DCT coefficients in R components, the experimental results are shown in Table 2. By comparing the results in Table 1 and Table 2, we find that the Markov-based features extracted from the DCT coefficients in R component perform better than the markov based features extracted from the DCT coefficients in Y component.

**Table 1: Experimental results using DCT coefficients from Y component with different T.**

| Threshold(dimension) | TP(%) | TN(%) | Accuracy(%) |
|---|---|---|---|
| T=2(100) | 98.43 | 98.45 | 98.44 |
| T=3(196) | 98.63 | 98.80 | 98.73 |
| T=4(324) | 98.66 | 98.95 | 98.81 |

**Table 2: Experimental results using DCT coefficients from R component with different T.**

| Threshold(dimension) | TP(%) | TN(%) | Accuracy(%) |
|---|---|---|---|
| T=2(100) | 99.27 | 98.99 | 99.13 |
| T=3(196) | 99.37 | 99.28 | 99.32 |
| T=4(324) | 99.30 | 99.36 | 99.33 |

On the other hand, both in the Table 1 and Table 2, we provide the performance of Markov-based features with three different T's, i.e., $T = 2, 3$ and 4, respectively. From these two table, we can notice that the performance of $T = 3$ is comparable with the performance of $T = 4$. However, when $T$ becomes much larger, the computational cost and the feature dimension increase sharply. To make a compromise between computational complexity and classification accuracy, we conclude that the Markov based features extracted from the DCT coefficients in R components with threshold $T = 3$ is the best choice for classify the real-scene images and recaptured images taken from printed pictures.

Other methods are also conducted on this generalized database. Table 3 shows the performance of other methods presented in literature and our proposed features extracted from the DCT coefficients in R component with threshold $T = 3$. The corresponding receiver operating characteristic (ROC) curves are given in Fig.2. Multi-scale Wavelet Statistics (MSWS) and Local Binary Pattern(LBP) were proposed in [6]. LBP was firstly presented by [11] to classify texture. Mode-based First Digit Features were firstly proposed in [12] and Yin et al. used these features to detect recaptured images taken from LCD screen in [7]. These three features are claimed to perform very well in classifying real-scene images and recaptured images. From the results in Table 3, we can notice that our Markov-based features extracted from the DCT coefficients array in R component with $T = 3$ outperform these state-of-art methods.

## 3.2 Experiments on a Specialized Database

As describe in section 3.1, in the process of establishing the generalized image database, the cameras used for photographic copying are not the same as the camera which take the real-scene images and the content of recaptured images are not the same with the content of real-scene images. The reason why we adopt this experimental setup is as following. In the reality, people always only have the printed pictures, such as the pictures printed on magazine, therefore people can not know the camera model and have the original real-scene images corresponding to the printed pictures. Although these setups meet the reality, the classification results of our proposed features may be doubted that the features works on identifying source cameras or classifying the content of images.

In order to remove this doubt, we establish another spe-

Table 3: **Experimental performance comparison for the proposed method.**

| Features(dimension) | TP(%) | TN(%) | Accuracy(%) |
|---|---|---|---|
| MSWS(54) | 96.04 | 96.16 | 96.10 |
| LBP(80) | 97.20 | 97.36 | 97.28 |
| MBFDF(180) | 96.57 | 98.15 | 97.36 |
| The proposed(T=3)(196) | 99.37 | 99.28 | 99.32 |

**Table 4: Experimental results conducted on the specialized database.**

| Features(dimension) | TP(%) | TN(%) | Accuracy(%) |
|---|---|---|---|
| MSWS(54) | 97.45 | 96.0 | 97.03 |
| LBP(80) | 95.70 | 97.15 | 96.43 |
| MBFDF(180) | 93.90 | 92.35 | 93.13 |
| The proposed(T=3)(196) | 97.75 | 96.60 | 97.18 |

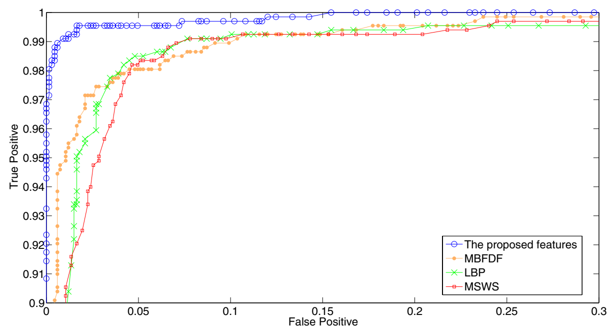


**Figure 2: The ROC curves**

cialized image database. We firstly use two cameras to take 600 real-scene images(300 images for each camera). Then we develop these images, consequently we get 600 printed pictures. At last, by using the same cameras to recapture these printed pictures, we obtain 600 recaptured images. For this new image database, the recaptured images have the same content with the real-scene images and the recapturing devices are the same cameras as original. The proposed features extracted from the DCT coefficients in R component with $T = 3$ are conducted on this image database and the classification results are shown in Table 4. The performance of other methods mentioned in Section 3.1 are also presented in Table 4. From Table 4, we can notice that the accuracy of these methods decline a little bit. This decline may caused by the flowing two reasons. Firstly, the printed pictures on magazines are poorer than the pictures printed on 4R glossy photo papers. Therefore, the quality of the recaptured images in this specialized image database may be higher than the recaptured images in the generalized image database. Secondly, there are less images used for training the classifier when all the methods are conducted on the specialized image database. Consequently, the classifier may perform worse. However, the performance of our proposed features is still better than others, which indicates that our proposed methods are independent of image contents and camera models.

# 4. CONCLUSIONS AND FUTURE WORK

In this work, we focus on detecting recaptured images taken from printed pictures. Based on our observation and analysis, the Markov based features are proposed to separate recaptured images from real-scene images. The Markov based features are extracted from the DCT coefficients. Two kinds of DCT coefficients, i.e., the quantized DCT coefficients in Y component and the DCT coefficients obtained by applying $8 \times 8$ block discrete cosine transform to the R component(the DCT coefficients in R component), and different threshold T are considered. Experiments conducted on a generalized image database including 3994 real-scene images and 3994 recaptured images taken from printed pictures show that using the DCT coefficients in R component with threshold $T = 3$ is the best choice. Comparison results show the proposed features work excellently in separating recaptured images from real-scene images and outperform the state-of-art methods.

In thefuture, we will further consider other kinds of photographic copying introduced in Fig.1. As different kinds of photographic copying may results in similar disturbance on the statistics of the images, we will extend our proposed methods to detect other kinds of photographic copying.

# 5. REFERENCES


[1] C. Holden, "Rare-tiger photo flap makes fur fly in China," *Science*, vol. 318, no. 5852, p. 893, 2007.

[2] D. Ngo, "Vietnamese security firm:Your face is easy to fake.[online]," available: http://news.cnet.com/8301-17938_105-10110987-1.html.

[3] H. Yu, T. T. Ng, and Q. Sun, "Recaptured photo detection using specularity distribution," in *Proc. of IEEE Int. Conf. on Image Processing*, San Diego, California, USA, Oct. 12-15, 2008.

[4] X. Gao, T. T. Ng, B. Qiu, and S. F. Chang, "Single view recaptured image detection based on physics based features," in *Proc. of IEEE Int. Conf. on Multimedia and Expo*, Singapore, July 19-23, 2010.

[5] J. Bai, T. T. Ng, X. Gao, and Y. Q. Shi, "Is physics based liveness detection truly possible with a single image?" in *Proc. of 2010 IEEE International Symposium on Circuits and Systems*, Paris, France, 2008.

[6] H. Cao and A. C. Kot, "Identification of recaptured photographs on LCD screen," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Dallas, Texas, USA, Mar. 14-19, 2010.

[7] J. Yin and Y. Fang, "Digital image forensics for photographic copying," in *Proceeding of SPIE Media Watermarking, Security, and Forensics 2012*, Burlingame, California, USA, Jan. 24-26 2012.

[8] Y. Q. Shi, C. Chen, and W. Chen, "A markov process based approach to effective attacking JPEG steganography," in *Proc. of Information Hiding Workshop*, Old Town Alexandria, Virginia, USA, July10-12, 2006.

[9] A. Leon-Garcia, *Probability,Statistics,and Random Processes for Electrical Engineering.* 3Edition, Prentice Hall, 2008.

[10] C. C. Chang and C. J. Lin, "LIBSVM – A library for support vector machines," available: http://www.csie. ntu. edu.tw/~cjlin/libsvm.

[11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.

[12] B. Li, Y. Q. Shi, and J. Huang, "Detecting doubly compressed JPEG images by using mode based first digit features," in *Proc. of IEEE Int. Workshop on Multimedia Signal Processing*, 2008, pp. 730–735.