Recaptured Images Forensics Based On Color Moments and DCT Coefficients Features

Rongrong Ni, Yao Zhao, and Xiaobo Zhai

Institute of Information Science
and Beijing Key Laboratory of Advanced Information Science and Network Technology
Beijing Jiaotong University, Beijing 100044, China
rrni@bjtu.edu.cn; yzhao@bjtu.edu.cn

ABSTRACT. *With the development of multimedia technology and digital devices, it is increasingly easier to photograph a high quality image. Due to the facility of capture process, recapture phenomenon becomes popular, which is harmful sometimes. In this paper, an effective recaptured image forensics algorithm through color moments and DCT coefficients features is proposed. Central moments in chromatic space are analyzed to verify the effectiveness in the aspect of detecting recaptured images. The mean, the standard deviation, and the third root of the skewness form the principal features. Furthermore, mode based first digit features of DCT coefficients in both luminance component and chrominance component are presented to distinguish the recaptured images from the real-scene images. Finally, these two kinds of features are combined to improve the detection performance. Experimental results demonstrate the proposed method achieves better performance in terms of accuracy.*
Keywords: Recaptured image forensics, Color moments, Mode based first digit features.

1. **Introduction.** With the development of multimedia technology and hardware devices, it is more and more easier to capture and process multimedia. Portable image and video shot devices are benefit for media capture. Furthermore, many software tools help us edit or compose the multimedia sources. As a result, forged or doctored multimedia are easily obtained, which raises an important issue of content security. The wide use of the Internet aggravates this phenomenon. Forensics technology analyzes the inherent characteristics of the multimedia, and reveals the abnormal differences to identify the integrity of the content. Many schemes have been proposed to detect the editing operations, such as splicing operation[1], JPEG compression[2, 3], contrast enhancement[4], and so on.

Image recapture detection is one kind of forensic techniques, which is to discriminate between the real-scene images and the recaptured images. Today, people can use a camera to photograph a real-scene image. When the image is shown on a screen or printed on a paper, we can shoot the image to produce a recaptured image.

Recapture events have also raised all kinds of security problems. Nowadays, face recognition system serves as an applicable access control system, where legal customs will pass the identification system. However, many face recognition authentication mechanisms allow anyone to log into a system easily with a photo of the legit owner, for example the laptops' authentication system. As a result, the illegal person can also pass through the control system by using one photo of the legal customs even at the highest authentication level[5]. It is an obvious menace to the security system. Another famous example

happened in 2007. One peasant claimed that he found the South China Tiger which was a kind of endangered animal. He provided some photos as evidences, as shown in Fig.1. His claim abstracted the scientists and governments all around of the world. But These photos were detected as rephotographed from a new year picture containing a tiger, as shown in the background of Fig.1.

FIGURE 1. Recaptured photo of South China Tiger

A tampered image can generally be detected by current forensic techniques because tampering process damages the consistency of camera characteristics. Whereas, if a tampered image, shown on a screen or printed on a paper, is re-photography by a camera, it may pass the forensic system. The reason is that the tampered image is really captured by a camera so that it presents most features of physical devices.

In addition, in many supermarkets automatic depositing bag cabinets are provided to store custom's bags. This system will give out a bar code scrip after the bag is put inside. When we want to take out the bag, this bar code is captured by a scanner and the corresponding door is opened. However, if someone takes a photo of the bar code, then shows the photo to the automatic deposit system, the door can also be opened. As shown in Fig.2, a bar code already captured by a cell phone can open the door of the cabinet.

FIGURE 2. Security issue in supermarkets

In this paper, we focus on the detection of recaptured images shown on LCD (liquid crystal display) screen, and propose an effective recaptured images detection algorithm based on color moments and DCT coefficients features. Because of the differences of light and environments, recaptured images exhibit the color diversity. An image is converted from RGB space to HSV space, then the central moments of chromatic spaces are calculated to form a color moment feature. From the aspect of compression, the recaptured

images undergo double JPEG compression. A mode based first digit feature is applied both in luminance component and in chrominance component to generate an effective detection characteristic. Finally, the combined feature is constituted to input into a SVM classifier. Experiments on three datasets verify the effectiveness and advantages of our method.

The rest of this paper is organized as follows: Related works are introduced briefly in Section 2, the proposed algorithm and the effectiveness analysis are given in Section 3. Section 4 provides the experiments, and followed by conclusions in Section 5.

2. **Related works.** Recently researches have shown great interests in detecting the recaptured images. The recaptured image reflects the characteristics of the media on which the original image is shown. Ng et al. analyzed the specular component of a recaptured image, which revealed the micro-texture and mesostructure of the printing paper [6]. The distinctive feature of a real-scene image and the corresponding recaptured one is in form of the distribution of the specular ratio's gradient. The specularity was further utilized to identify live faces and recaptured faces in [7]. The specular gradient histogram was quantified to reliably detect the presence of a display medium. In [8], a general recapture model was presented and a set of physics-based features were analyzed to recognize the recaptured images, which consisted of the contextual background information, the distribution of specularity, the image gradient, color histogram, local contrast, chromaticity feature, and a blurriness measure.

Cao et al. pointed out that the finely recaptured images were threats for both image forensic system and human eyes [9]. A set of statistical features were proposed including local binary pattern (LBP) texture feature, multi-scale wavelet statistics measuring the loss of fine details, color features reflecting the color anomalies. These features were used for recaptured image forensics in an image database which was set up with finely controllable settings. Kose and Dugelay proposed a method based on contrast and texture characteristics to identify real face images and recaptured face images [10]. Considering the robustness of detection, a rotation invariant local binary pattern variance (LBPV) was used for face spoofing detection.

Yin et al. [11] used the noise features and the mode based first digital features to detect the recaptured images. The noise feature was extracted by using three different DWTs (discrete wavelet transforms) to maximize the differences. A mode based first digital feature in luminance component was used to discriminate the recaptured images.

Recapture process is inside the acquisition chain that was modeled in [12]. The acquisition stage of an image was identified by constructing dictionaries of edge profiles. The method is valid for both recapture detection and chain identification applications.

3. **Proposed features and effectiveness analysis.** Image acquisition and storage are two basic aspects to obtain a digital image. For image acquisition, since the shooting situations of the recaptured images and the real-scene images will never be same, the real-scene images and the corresponding recaptured ones manifest different color characteristics which are caused by the type of light, the direction of light and the light intensity and so forth. For image storage, JPEG is a popular image format which is applied automatically during photography. The recaptured images suffer double JPEG compression due to the second camera shooting, which will change the DCT coefficients. Based on the above facts, color feature and DCT coefficients feature are combined to differentiate the recaptured images from the real-scene ones.

3.1. **Color feature selection and the effectiveness.** Color is a widely used feature for image representation. Color histogram and cumulative color histogram are effective feature descriptions for images. These features contain the complete color distribution, so the dimension tends to be high and increases computing complexity. Therefore, low dimension feature is a good choice to measure color distribution. It is imperative that low dimension features have the potential to characterize the color images.

Instead of using high dimensional color features, we only consider the color moments as effective dominant features. Color moment was proposed by M. Stricker and M. Orengo[13], and has already been used to identify computer generated images [14]. If the color distribution is regarded as a kind of probability distribution, the central moments can represent the color images.

If the value of the ith color component at the jth image pixel is p_{ij}. The mean in the ith color channel is given in Eq.(1), where N is the number of pixels in the image. The k-order central moment in the ith color channel is formulated as Eq.(2). The mean, the variance and the skewness are the first three central moments.

$$\mu_i = \frac{1}{N} \sum_{j=1}^{N} p_{ij} \tag{1}$$

$$\mu_{ik} = \frac{1}{N} \sum_{j=1}^{N} (p_{ij} - \mu_i)^k, k = 2, 3, 4... \tag{2}$$

Correspondingly, the standard deviation and the third root of the skewness in the ith color channel are described as follows,

The standard deviation is:

$$\sigma_i = (\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - \mu_i)^2)^{1/2} \tag{3}$$

The third root of the skewness is:

$$s_i = (\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - \mu_i)^3)^{1/3} \tag{4}$$

Hue and saturation is more consistent with human perception in terms of the description of an image. In this paper, we use HSV model as the color space. The first step is to convert an image from RGB space to HSV space, which means Hue, Saturation and Value component respectively. Then the first three moments in each color space are calculated. Finally, the mean, the standard deviation and the third root of the skewness of each color space are connected to form a 9-dimensional feature.

This color feature is described below:

$$feature_{color} = [\mu_H, \mu_S, \mu_V, \sigma_H, \sigma_S, \sigma_V, s_H, s_S, s_V] \tag{5}$$

This feature exhibits different characteristics of the real-scene images and the recaptured images, which is proved by the experiments. Two datasets are used in this section, which are interpreted in Section 4.1 in details. Dataset 1 is constructed by ourselves including 200 pairs of images; and dataset 2 includes 800 pairs of images provided by Ng et al.

In Fig.3, 200 pairs of images from dataset 1 is considered. $feature_{color}$ of each image is obtained and the average value is calculated. The red line with circle marks is the average color feature of 200 recaptured images; while the blue line with star marks manifests the average color feature of 200 real-scene images.

Fig.4 also shows the effectiveness of the color feature. Here, 800 pairs of images from dataset 2 is involved. The red line with circle marks is the average color feature of 800 recaptured images; while the blue line with star marks manifests the average color feature of 800 real-scene images.

From these two figures, we observe that the first 3 elements in $feature_{color}$ of the recaptured images are bigger than that of the real-scene ones, and the last 6 elements in $feature_{color}$ of the recaptured images are smaller than that of the real-scene ones. So $feature_{color}$ is a suitable characteristic to distinguish recaptured images from the real-scene ones.
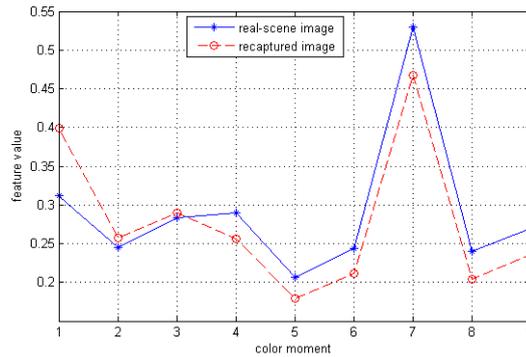

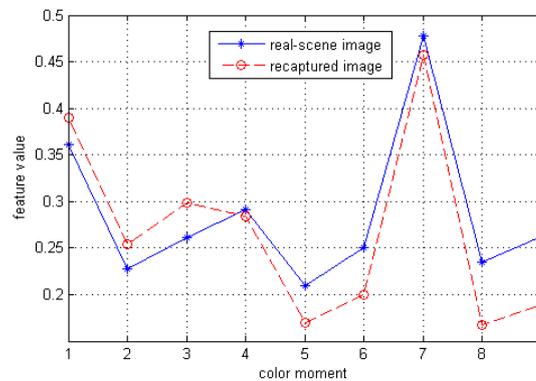FIGURE 3. The average value of color feature of 200 pairs of images from dataset 1

FIGURE 4. The average value of color feature of 800 pairs of images from dataset 2

3.2. **DCT coefficients feature and the effectiveness.** JPEG format is a kind of popular image format, which is based on Discrete Cosine Transform. When the real scene is shot by using a digital camera, the acquired digital image undergos JPEG compression once. While the image shown on certain media is captured again, the digital image is compressed for the second time. From the perspective of image acquisition, the recaptured images manifest some features of double JPEG compression.

There are several methods aiming at detecting double JPEG compression [15, 16]. Yin et al. [11] used $MBFDF$(mode based first digital feature) [16] to identify the recaptured images from the real-scene images. However, they only take into account the luminance component. As mentioned above, color component is an efficient feature for detecting the

recaptured images. We consider both color component and luminance component, and analyze the effectiveness in this paper.

In each 8×8 image block, DCT coefficients are ranked in Zigzag order, shown in Fig.5. There are one DC coefficient and 63 AC coefficients. Mode means the AC coefficient position in the Zigzag scanned order. So, mode changes from 1 to 63. In Fig.5, the positions in green are the first 5 modes.

The $MBFDF^{\phi}$ which considers both color component and luminance component is defined as Eq.(6):

$$P_i^{\phi}(d) = \frac{N_i^{\phi}(d)}{\sum_{m=1}^{9} N_i^{\phi}(m)}, d \in \{1, 2, 3, ..., 9\} \tag{6}$$

where $\phi \in \{Y, C_r, C_b\}$, and i represents one of the 63 modes in Zigzag order. $N_i^{\phi}(m)$ denotes the number of all the non-zero coefficients, whose first digit is m, at the ith mode.

FIGURE 5. DCT coefficient position and Zigzag scan order

For a JPEG image, the compressive bit stream contains luminance component and chrominance component. We can obtain the quantized DCT coefficients of Y component and Cr component by using JPEG Toolbox[17] based on IJG's (Independent JPEG Group) code library. Then, $MBFDF^Y$ is calculated from Y component based on Eq.(6), and $MBFDF^{Cr}$ is calculated from Cr component. If the first 5 modes are involved, we can obtain a 9*5*2=90 dimension feature.

Fig.6 shows a pair of images from dataset 1. Fig.7 gives the probabilities of the first digits of quantized DCT coefficients at the 2th and the 3th mode in Y component. The figures manifest the differences between the real-scene image and the recaptured image. The probabilities of the first digits extracted from Cr component are shown in Fig.8, from which the differences are more obvious and effective.
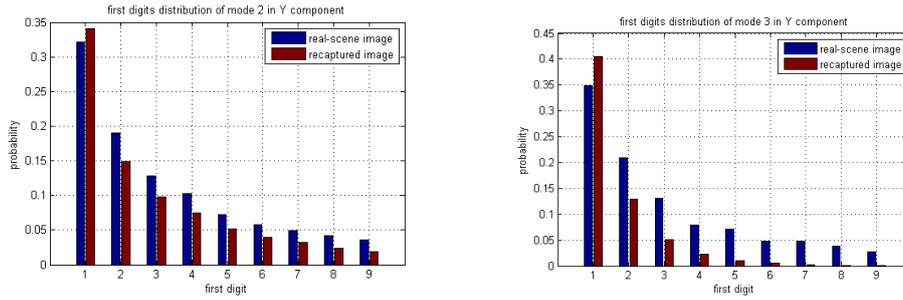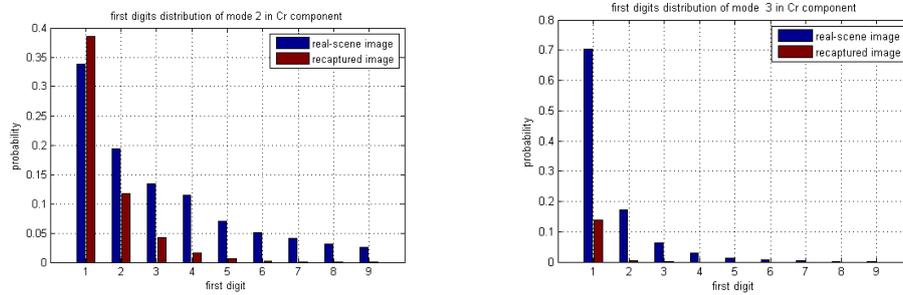

(a) The real-scene image   (b) The corresponding recaptured image

FIGURE 6. The real-scene image and corresponding recaptured one from dataset1

(a) The probability of mode 2 in Y component

(b) The probability of mode 3 in Y component

FIGURE 7. First digits distribution of mode 2 and mode 3 in Y component for the real-scene image and the corresponding recaptured one randomly chosen from dataset1

(a) The probability of mode 2 in Cr component

(b) The probability of mode 3 in Cr component

FIGURE 8. First digits distribution of mode 2 and mode 3 in Cr component for the real-scene image and the corresponding recaptured one randomly chosen from dataset1

Fig.9 shows a pair of images from dataset 2. Fig.10 gives the probabilities of the first digits of quantized DCT coefficients at the 2th and the 3th mode in Y component. The figures also reveal the differences between the real-scene image and the recaptured image. The probabilities of the first digits extracted from Cr component are shown in Fig.11, from which the differences can be caught more effectively.

(a) The real-scene image

(b) The corresponding recaptured image

FIGURE 9. The real-scene image and corresponding recaptured one from dataset2

From the above analysis, $MBFDF^{Y}$ and $MBFDF^{Cr}$ are effective features for detecting the recaptured images.

(a) The probability of mode 2 in Y compo- (b) The probability of mode 3 in Y com-
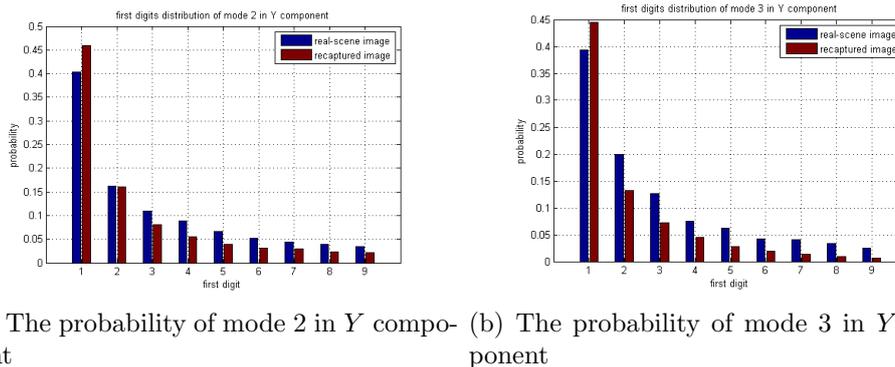nent   ponent

FIGURE 10. First digits distribution of mode 2 and mode 3 in Y component
for the real-scene image and the corresponding recaptured one randomly
chosen from dataset2

(a) The probability of mode 2 in Cr com- (b) The probability of mode 3 in Cr com-
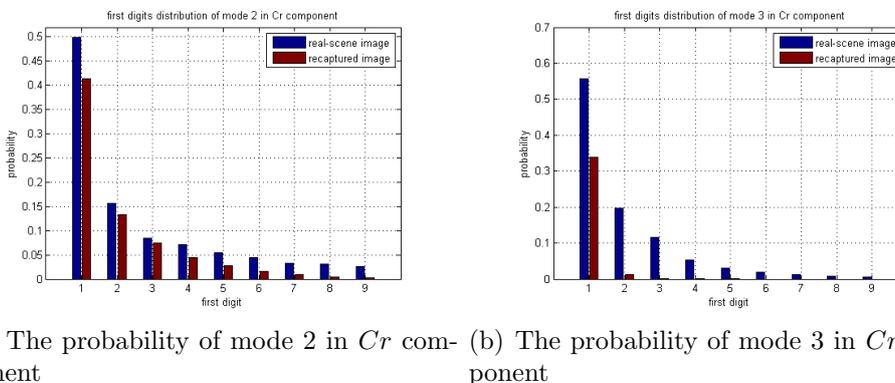ponent   ponent

FIGURE 11. First digits distribution of mode 2 and mode 3 in Cr com-
ponent for the real-scene image and the corresponding recaptured one ran-
domly chosen from dataset2

4. **Experiments.** In our research, the problem is to classify the real-scene images and
the recaptured images into two classes. Because SVM (support vector machine) is a good
classifier for this purpose, it is suitable to detect the recaptured images. After the features
described in the above section are extracted, SVM classifier is applied to decide whether
an image is original or recaptured.

4.1. **Dataset description.** In this paper, we use three datasets containing real-scene
images and recaptured images. Dataset 1 is constructed by ourselves. The real-scene
images are shot by using Single Lens Reflex cameras: Cannon EOS 500D and Nikon
D600, respectively. After the images are shown on LCD screen, we use these two cameras
to recapture all the images. The real-scene images are 200, including buildings, people,
natural scene, animals, plants and so on. Thus, the recaptured images are 400. In our
experiments, in order to balance the number of two classes of images, we use all 200
real-scene images, while choose 200 recaptured images.

Dataset 2 is provided by I^2R (Institute for Infocomm Research, Singapore) [8], which
includes 800 real-scene images and 800 recaptured ones. While the images are with small
size, for example, some images are smaller than 200*300 pixels.

In order to verify the effectiveness of the proposed algorithm in a larger dataset, we
construct Dataset 3 which includes 636 real-scene images and 636 recaptured images, all

TABLE 1. The experimental results on dataset 1

Features	Accuracy	Recall	Precision
LBP[9]	92.98	89.10	96.19
Color moments(CM)	85.07	82.16	88.75
$MBFDF^Y$	91.89	94.56	88.41
$MBFDF^{Cr}$	92.25	96.30	89.92
CM+$MBFDF^Y$+$MBFDF^{Cr}$	93.36	96.85	90.08

TABLE 2. The experimental results on dataset 2

Features	Accuracy	Recall	Precision
LBP[9]	93.47	89.90	95.24
Color moments(CM)	83.24	80.19	86.45
$MBFDF^Y$	90.67	93.62	87.93
$MBFDF^{Cr}$	92.78	95.16	89.29
CM+$MBFDF^Y$+$MBFDF^{Cr}$	93.86	96.75	90.46

with high quality and big size. The 636 real-scene images are captured by Cannon EOS 500D and Nikon D600, whose size is larger than 3456*2304 pixels. In the 636 recaptured images, there are 384 images provided by Sun Yat-sen University [11] and 252 images recaptured at the above real-scene images.

4.2. **Experimental results.** Recaptured images are regarded as the positive class denoted as P, while real-scene images as the negative class denoted as N. TP represents the recaptured images are classified as the recaptured ones. TN means the real-scene images are identified as the real-scene ones. FP represents the real-scene images are classified as the recaptured images. FN represents the recaptured images are identified as the real-scene ones.

Recall, Precision and Accuracy are defined as follows:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

For dataset 1 and dataset 2, we randomly select 2/3 images to train the SVM, and the rest 1/3 are used for test. In order to eliminate the randomness of the results, we repeat the process 10 times, and get the average as the results. The LIBSVM toolbox[18] is used, and the non-linear radial basis function(RBF) kernel is chosen in the experiments. We compare the proposed methods with the LBP approach in [9].

Table 1 shows the results on dataset 1, and Table 2 shows the results on dataset 2. From the tables, we note that in dataset 1 the average accuracy of color moments(CM), $MBFDF^Y$ and $MBFDF^{Cr}$ can reach 85.07%, 91.89%, and 92.25%, respectively. The accuracy of the combined feature reaches 93.36%, which is better than LBP. When dataset 2 is used, the average accuracy of color moments(CM), $MBFDF^Y$ and $MBFDF^{Cr}$ can reach 83.24%, 90.67%, and 92.78%, respectively. The accuracy of the combined feature reaches 93.36%, which also outperforms LBP.

TABLE 3. The experimental results on dataset 3

Features	Accuracy	Recall	Precision
LBP[9]	96.50	96.76	96.23
Color moments(CM)	77.10	75.72	79.77
$MBFDF^Y$	90.67	95.00	85.85
$MBFDF^{Cr}$	92.77	93.78	91.73
CM+$MBFDF^Y$+$MBFDF^{Cr}$	96.70	97.24	96.12

For dataset 3, we randomly select 5/6 images as training set, and the rest 1/6 are used for test. In order to eliminate the randomness of the results, we repeat the process 10 times, and get the average as the results. Table 3 gives the performance results on this larger dataset. The average accuracy of color moments(CM), $MBFDF^Y$ and $MBFDF^{Cr}$ can reach 77.10%, 90.67%, and 92.77%, respectively. The accuracy of the combined feature reaches 96.70%.

5. **Conclusions.** In this paper, we presented an effective recaptured images forensics algorithm through color moments and DCT coefficients features. Because the re-photography process introduces some changes to color perception, the central moments of chromatic space are valid description to the color distribution which is regarded as probability distribution. We analyzed the first three moments in HSV space, i.e. the mean, the standard deviation, and the third root of the skewness, to verify their effectiveness. Since recaptured images may undergo double JPEG compression, mode based first digit features in both luminance component and chrominance component are applied to efficiently distinguish the recaptured images from the real-scene images. Experiments show that the color moments, $MBFDF^Y$ and $MBFDF^{Cr}$ are good characteristics, and the combined features outperform LBP feature.

REFERENCES

[1] O. M. AI-Qershi, and B. E. Khoo, Passive detection of copy-move forgery in digital images: State-of-the-art, *Forensic Science International*, vol. 231, no. 1-3, pp. 284-295, 2013.

[2] T. Bianchi, and A. Piva, Image forgery localization via block-grained analysis of JPEG artifacts, *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 3, pp. 1003-1017, 2012.

[3] G. Valenzise, M. Tagliasacchi, and S. Tubaro, Revealing the traces of JPEG compression anti-forensics, *IEEE Trans. on Information Forensics and Security*, vol. 8, no. 2, pp. 335-349, 2013.

[4] G. Cao, Y. Zhao, R. Ni, and X. Li, Contrast enhancement-based forensics in digital images, *IEEE Trans. on Information Forensics and Security*, vol. 9, no. 3, pp. 515-525, 2014.

[5] X. Tan, Y. Li, J. Liu, and L. Jiang, Face liveness detection from a single image with sparse low rank bilinear discriminative model, *Proc. of European Conference on Computer Vision*, Crete, Greece, pp. 504-517, 2010.

[6] H. Yu, T. T. Ng and Q. Sun, Recaptured Photo Detection Using Specularity Distribution, *Proc. of IEEE International Conference on Image Processing*, San Diego, California, USA, pp. 3140-3143, 2008.

[7] J. Bai, T. T. Ng, X. Gao and Y. Q. Shi, Is Physics-based Liveness Detection Truly Possible with a Single Image?, *Proc. of IEEE International Symposium on Circuits and Systems*, Paris, France, pp. 3425-3428, 2010.

[8] X. Gao, T. T. Ng, B. Qiu and S. F. Chang, Single-view Recaptured Image Detection Based on Physics-based Features, *Proc. of IEEE International Conference on Multimedia and Expo*, Singapore, pp. 1469-1474, 2010.

[9] H. Cao, and A. C. Kot, Identification of Recaptured Photographs on LCD Screens, *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, Texas, USA, pp. 1790-1793, 2010.

[10] N. Kose, and J. L. Dugelay, Classification of Captured and Recaptured Images to Detect Photograph Spoofing, *Proc. of the 1st International Conference on Informatics, Electronics and Vision*, Dhaka, Bangladesh, pp. 1027-1032, 2012.

[11] J. Yin, and Y. Fang, Digital image forensics for photographic copying, *Proc. of SPIE Electronic Imaging: Media watermarking, Security and Forensics*, San Jose, California, USA, pp. 8303F-1-8303F-7, 2012.

[12] T. Thongkamwitoon, H. Muammar, and P. L. Dragotti, Identification of image acquisition chains using a dictionary of edge profiles, *Proc. of European Signal Processing Conference*, Bucharest, Romania, pp. 1757-1761, 2012.

[13] M. Stricker, and M. Orengo, Similarity of color images, *Proc. of SPIE Storage and Retrieval for image and Video Databases Conference*, San Jose, California, USA, pp. 381-392, 1995.

[14] Y. Chen, Z. Li, M. Li, and W. Y. Ma, Automatic Classification of Photographs and Graphics, *Proc. of International Conference on Multimedia and Expo*, Toronto, Canada, vol. 9-12, pp. 973-976, 2006.

[15] A. C. Popescu, *Statistical Tools for Digital Image Forensics*, Ph.D. Thesis, Department of Computer Science, Dartmouth College, Hanover, New Hampshire, USA, 2004.

[16] B. Li, Y. Q. Shi, and J. Huang, Detecting doubly compressed JPEG images by using mode based first digit features, *Proc. of IEEE International Workshop on Multimedia Signal Processing*, Queensland, Australia, pp. 730-735, 2008.

[17] Phil Sallee, Matlab JPEG Toolbox, http://xdatasecurity.com/matlab-based/phil-sallee-matlab-jpeg-toolbox.htm.

[18] C. C. Chang, and C. J. Lin, LIBSVM : a library for support vector machines, http://www.csie.ntu.edu.tw/ cjlin/libsvm.