Color Histogram Based Image Forgery Detection for Copy-Move Attack

Bo Liu and Chi-Man Pun

Faculty of Science and Technology, University of Macau,
Av. Padre Tomas Pereira, Taipa, Macau, China
{mb05422, cmpun} @umac.mo

Abstract. Copy-move attack is a common form of image forgery which belongs to a research field in multimedia forensics. In this paper, a blind forensics method based on color histogram is proposed to detect such forgery in image. The test image is firstly resized and converted to HSV color space, and its color value is quantized. And the image is divided into fixed-size overlapped blocks and the feature vector based on histogram of quantized color value is constructed for every block. Then the pair blocks are found by sorting all vectors. Lastly all blocks in those pairs are clustered based on their spatial location and the system recognizes the duplicated regions by evaluating correlation between clusters. The experimental results demonstrate that the proposed approach can not only detect ordinary copy-move attack for various kinds of pictures, but also recognize the cloned regions which are incidentally changed by rotation and scale transformation.

Keywords: Digital image forensics, copy-move forgery, color histogram

1 Introduction

As the great development of digital photography and relevant post-processing technology, digital image forgery becomes easily in terms of operating thus may be improperly utilized in news photography in which any forgery is strictly prohibited or the other scenario, for instance, as an evidence in the court. Therefore, digital image forgery detection technique is needed. Forgery detection technique is therefore becoming more and more important. Watermarking [1] can be a tool to help detect improper modifications on image. In fact, this technique requires operation on original taken photo and cannot deal with source-unknown images. Many proposed methods [2] thus avoid utilizing watermarking and other signature scheme.

A widely used image forgery means is copy-move attack which means duplicating one area and then pasted onto other zones within an image. SIFT [3] has been used to find cloned objects [4] but it fails to deal with the situation where SIFT feature scarce area is copied and pasted. For example, copy a region of peace blue sky to cover a bird may not be detected by this method. Several methods search for correlation between duplicated areas by dividing image into overlapping blocks and then applying block-based quantized Discrete Cosine Transform (DCT) [5] to find such correlation. Considering the huge feature vector length, Li et al [6] use Singular Value Decomposition (SVD) to reduce feature yielding from Discrete Wavelet Transform (DWT). Zimba et al [7] reduce image dimension by considering only low frequency sub-band of Discrete Wavelet Transform (DWT) and then applying Principle Component Analysis-Eigen Value Decomposition (PCA-EVD) to reduce feature length. Also, Ghorbani et al [8] combine Discrete Wavelet Transform (DWT) and Discrete Cosine Tranform (DCT) to construct reduced feature. But these methods are not robust to rotation and scale duplicating operation. Adopting Fourier-Mellin Transform (FMT) and vector erosion filter [9] is robust to arbitrary rotating, slightly scaling and JPEG compression.

In this paper, the proposed color histogram based forgery detection approach is effective on various kinds of digital image while [4] only works with area that SIFT feature can be constructed. Also, our method deals with brightness manipulation and is robust to rotation and scale transformation.

2 Color Histogram Feature for Forgery **Detection**

The proposed method is mainly based on color histogram feature with which copy-move attack in a single digital color image can be detected because the duplicated regions share the same color feature. The reason why we choose HSV rather than RGB color space to attract feature is that the feature length of brightness (V) can be adjusted to make the algorithm robust to object illumination change which may be used by forger in order to make forgery more realistic. A simple illustration of the whole forgery detection process is: the first step consists of image size normalization and block feature construction, the second step involves block matching and clustering, the third one is devoted to cluster comparison and forgery detection. The overall framework of our proposed method is shown in Fig. 1.

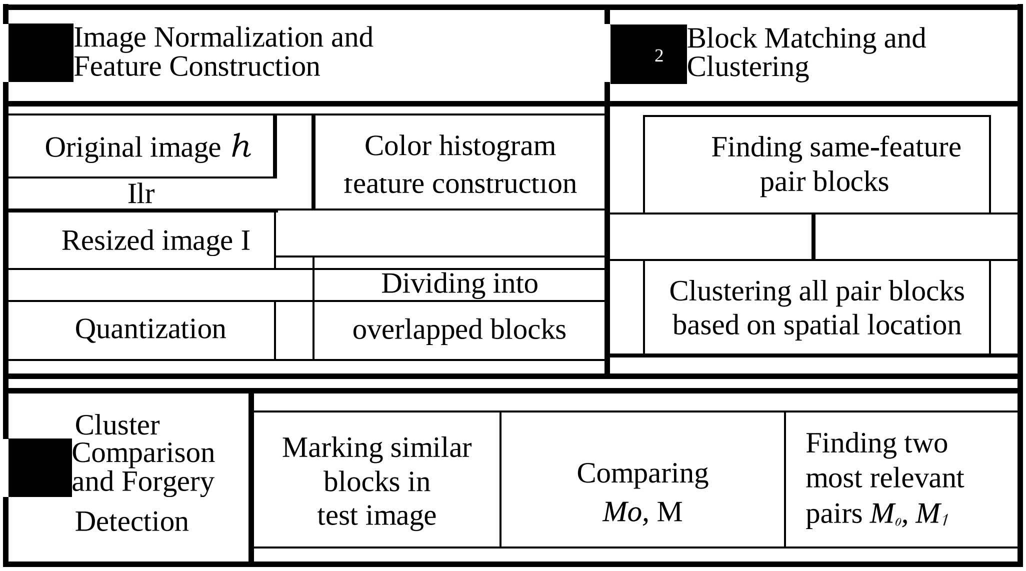
Image Normalization and Feature Construction
2
Block Matching and Clustering
Original image h
Ilr
Color histogram feature construction
Finding same-feature pair blocks
Resized image I
Dividing into overlapped blocks
Quantization
Clustering all pair blocks based on spatial location
Cluster Comparison and Forgery Detection
Marking similar blocks in test image
Comparing Mo, M
Finding two most relevant pairs M0, M1

Fig. 1. The overall framework of proposed method.

2.1 Image Normalization and Feature Construction

The test image I_t is initially resized to *M* by *N* pixels, denoted by *I*, and the hue, saturation and brightness value is quantized to I_H, I_s and I_v discrete levels whose value $(s_i,s2,...,s_{/s})$ and $(v_i,v_2,...,v_{/v})$ respectively. Then *I* is divided into *(M-b+1) X(N-b+1)* fixed-size overlapped blocks, where *b* is the side length (pixels) of each block. An initial ***0*** vector $P=(0,0,...,0)1H+IS+IV$ is constructed for every block. Note that the *P* is made up of three parts which relate to quantized hue, saturation and brightness value respectively. More specifically, one position in *P* corresponds to a HSV value in table 1.

Table 1. Correspondence between vector **P** and quantized HSV value.

	H				*S*				*V*				
P= (	0	0	...	0	0	0	...	0	0	0	...	0	)
	h./	*h2*	...	*hJH*	*si*	*s2*	. ..	*SIS*	*VI*	*V2*	. ..	*VIV*	

Then, for each position in vector *P*, the new value is the number of pixels whose value equal to its corresponding HSV quantized value in the third row of table 1. There are *(M-b+1) x(N-b+1)* feature vectors in total need to construct. And after this, it can establish a *[(M-b+1)x(N-b+1)]* by $(/_H+/_s+/_v)$ matrix ***Q*** to store the feature vector of each block for future processing.

2.2 Block Matching and Clustering

Considering the cloned region is consisted of some blocks, they can theoretically be detected by finding the same rows in ***Q.*** The system stores all these blocks in a matrix ***K*** which is, obviously, m by $(I_H+I_s+I_v)$, and m is the number of detected blocks. However, there are some matched pair blocks which share the same color feature but they are not artificially duplicated in experiment. These false matched blocks of a test image appear in the pure color region such as peace blue sky and overexposure area in which blocks' color is the same or changed so slightly that can hardly be distinguished as two different ones because of limited discrete value bins described previously. But notice that these false matched blocks are spatially random and independent, and therefore it can be solved by clustering detected blocks in ***K.***

K-means clustering is a method of cluster analysis and it aims to partition *m* observations into *k* clusters in which each observations belongs to the cluster with the minimum within-cluster sum of squares (WCSS). Given a set of observations $(x_i, x_2, x_3,..., x_n)$, where x_i is a d-dimensional real vector. K-means cluster the set into *k*

n) parts *S={SI, S2, S3,..• Sk }* so as to minimize WCSS. The blocks in ***K*** are segmented into *k* clusters $S=\{S_i,S_2,...,S_k\}$ by k-means algorithm according to their spatial location.

2.3 Cluster Comparison and Forgery Detection

This step evaluates the relevance of every two clusters in S which was introduced by previous step. The most relevant pair in S is the one that contains the largest number of matched blocks. Name this most relevant pair MO, and similarly the pair containing the second largest number is marked as Ml and so on. Assume there is only one region is copied and pasted from other parts of the image, the system will show the cloned regions if

$$= Miro < P > P^{E} \left(^{CI,\ 1}\right) \qquad (^{1})$$

The parameter p is set manually and in our experiment: p=0.5.

3 Experimental Results

In this section, we evaluate our proposed method with a small dataset. All pictures are resized to around 600 by 400 pixels. Setting proper values for parameters will be discussed before showing detecting results.

3.1 Setting Parameters

In our proposed method, there are five important parameters: the size of overlapped blocks *b*, bin number $/_H$, I_S, I_v and the number of cluster *k*.

In order to improve performance, some constrains of parameters must be taken into consideration. For parameter *b*, on the one hand, a bigger value of *b* will increase program's efficiency because the total number of blocks in test image is inversely proportional to the value of *b*. But it cannot exceed the size of cloned object. However, such size varies from picture to picture and thus cannot be predicted in advance. A safe and reasonable way is to set *b* smaller. On the other hand, if *b* is set too small, the feature vector constructed from each block may be short and then produce too many matched blocks. In such circumstance, more clusters *k* is necessary to eliminate those "clutter" matched pairs. For parameter I_v, it should be set smaller than ***hi*** and ***Is*** to fight against brightness change. After our experiment, the recommended parameter values are listed in table 2.

Table 2. Recommanded parameter values.

III	***IS***	***Iv***	***b***	***k***
32	32	8	5	45

3.2 Detection Results and Discussion

This part initially presents the detection result of basic copy-move attack, which means we just duplicate an object then parallel translate and paste onto other part within the image. This will verify the effectiveness of our method. The Fig. 2 below consists of three images, which are, from left to right, original taken picture in JPEG format, tampered image and the detection result. Two different colors indicate the cloned regions connected by black lines. In order to show that this approach fits for various size of cloned object, different images are also revealed in Fig. 3. But there is one thing needs to point out. From the experimental results presented, it can easily be observed that only part of cloned area is marked and this is because of the cluster configuration issue. Concretely speaking, the number of cluster *k*, as a parameter, is set manually in advance. If *k* is set large, cloned area will be made up of more than two clusters, however the algorithm will only highlight two clusters. Despite only part of region will be shown, it is enough to distinguish forgeries.

Fig. 2.

2. our algorithm: (a) original picture, (b) tampered picture, (c) detectio

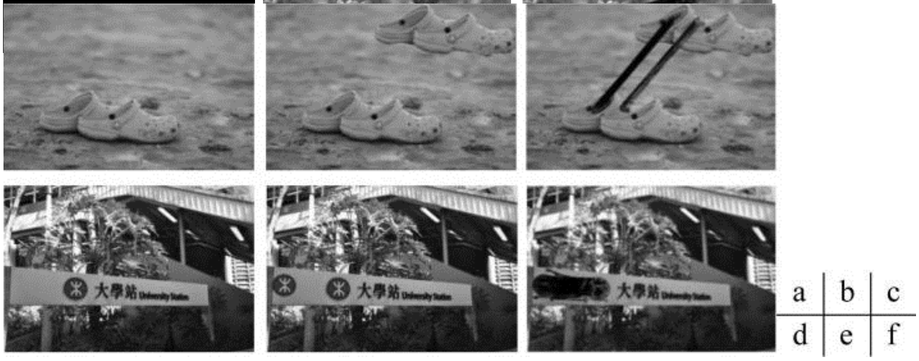

Fig. 3. Detection of cloned object in different size: (a) (d) original picture, (b) (e) tampered picture, (c) (f) detection results.

3.3 Robust Against Geometrical Changes

Then the situation when the duplicated object is deformed by rotation and scale operation is presented. Methods proposed by [6][7][8] can hardly detect copy-move forgery when these geometrical changes are applied. But our method is applicable and effective as shown in Fig. 4.

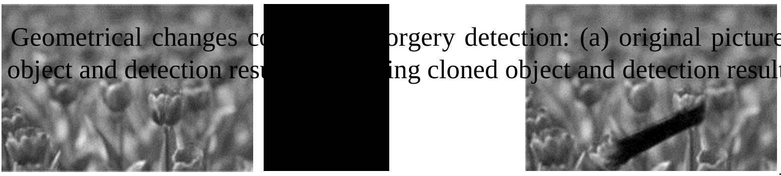

Fig. 4. Geometrical changes c [illegible] orgery detection: (a) original picture, (b) enlarging cloned object and detection res [illegible] ing cloned object and detection results.

4 Conclusions

In this paper, the problem of copy-move forgery detection is discussed. The proposed algorithm is based on HSV color histogram to deal with such attack and shows good performance and most importantly, can deal with rotation and scale transformation. The feature work includes research on automatic parameters setting, and integration with other forensics techniques will be investigated.

References

1. M. Barni and F. Bartolini, Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications. NewYork: Marcel Dekker, 2004.
2. J. A. Redi, W. Taktak, and J.-L. Dugelay, "Digital image forensics: A booklet for beginners," Multimedia Tools Applicat., vol. 51, no. 1, pp.133-162, 2011.
3. David G. Lowe, "Object recognition from local scale-invariant features," International Conference on Computer Vision, Corfu, Greece (September 1999), pp. 1150-1157.
4. Amerini, I.; Ballan, L.; Caldelli, R.; Del Bimbo, A.; Serra, G.; , "A SIFT-Based Forensic Method for Copy—Move Attack Detection and Transformation Recovery," Information Forensics and Security, IEEE Transactions on , vol.6, no.3, pp.1099-1110, Sept. 2011.
5. J. Fridrich, D. Soukal, and J. Lukas, "Detection of Copy-Move Forgery in Digital Images," in Proceedings of Digital Forensic Research Workshop, August 2003.
6. Guohui Li; Qiong Wu; Dan Tu; Shaojie Sun; , "A Sorted Neighborhood Approach for Detecting Duplicated Regions in Image Forgeries Based on DWT and SVD," Multimedia and Expo, 2007 IEEE International Conference on , vol., no., pp.1750-1753, 2-5 July 2007.
7. M. Zimba, S. Xingming , "DWT-PCA (EVD) Based Copy-move Image Forgery Detection", International Journal of Digital Content Technology and its Applications, pp. 251-258, January 2011.
8. Ghorbani, M.; Firouzmand, M.; Faraahi, A.; , "DWT-DCT (QCD) based copy-move image forgery detection," Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on , vol., no., pp.1-4, 16-18 June 2011.
9. Weihai Li; Nenghai Yu; , "Rotation robust detection of copy-move forgery," Image Processing (ICIP), 2010 17th IEEE International Conference on , vol., no., pp.2113-2116, 26-29 Sept. 2010.