An Efficient Clustering Technique for Copy-Paste Attack Detection

N. Chaitawittanun, M. Munlin

Abstract**—**Due to rapid advancement of powerful image processing software, digital images are easy to manipulate and modify by ordinary people. Lots of digital images are edited for a specific purpose and more difficult to distinguish form their original ones. We propose a clustering method to detect a copy-move image forgery of JPEG, BMP, TIFF, and PNG. The process starts with reducing the color of the photos. Then, we use the clustering technique to divide information of measuring data by Hausdorff Distance. The result shows that the purposed methods is capable of inspecting the image file and correctly identify the forgery.

Keywords**—**Image detection, forgery image, copy-paste.

I. Introduction

PHOTOGRAPHY is one of the most popular and interesting activities that can be performed anywhere. The major equipment of taking a photo is a digital camera which is convenient, inexpensive, and easy to use. In addition, it can save the images and instantly display them [1]. That is the reason why a photo from digital cameras is popularly used in many media, such as newspapers, magazines, social network, even including crime scene evidences [2]. These photos may contain important events and be used as evidence. Nowadays, there is a doubt whether the pictures have been changed or not [3]. Retouching photos is now harmfully cultural competence spreading all over the internet. Celebrities, actors, politicians, even civilians can be the victim of retouching [4]. Tampering images might lead harmfully to misunderstanding or misleading the truth which the suspect reputation of people in the photos.

The image forgery detection is classified into active and passive [5]. Active process uses digital watermark to examine the fake images as shown in Fig. 1. Hiding information into the image before using can be used to examine the history of that image [6]. Nevertheless, this technique also has limitations such as the user has to know how to embed the secret information onto the image. This technique is inappropriate and difficult to inspect pictures.

In contrast, the passive technique does not embed information into a picture. This technique is much more appropriate and easier. Many researchers have focused on passive image forgery detection. Fridrich [7] has developed techniques of overlapping block and DCT which extract feature of the images, then compare the similarity of block. Popsecu and Farid [8] use PCA to reduce the dimension of the block. Farid [8] suggest the idea of using vision of human being to examine shadow of objects, reflection of objects and distortion of objects. The image forgery detection methods that have been developed by most researchers are feature extraction by using various techniques such as DCT and PCA.

There are two types of popular techniques of attack; Copy-paste and Image Slicing. In Copy-paste image, forgery takes some parts of the image and paste onto another part of the same image as shown in Fig. 3.

Image Splicing technique, on the other hand, takes some parts from two or more images to create a new image and change the look of the original image [7] as shown in Fig. 4.

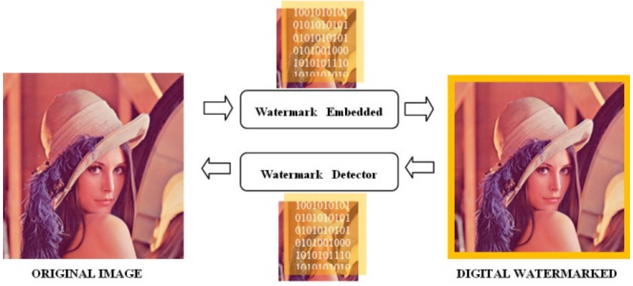


Fig. 1 Digital watermark process

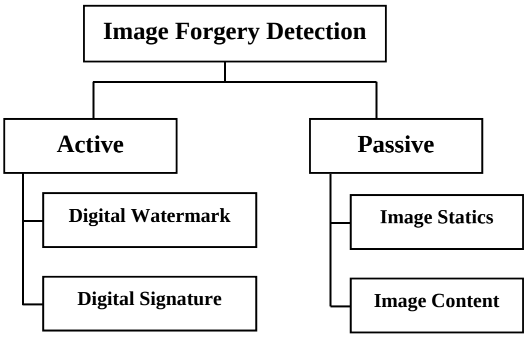


Fig. 2 Classification of image forgery detection

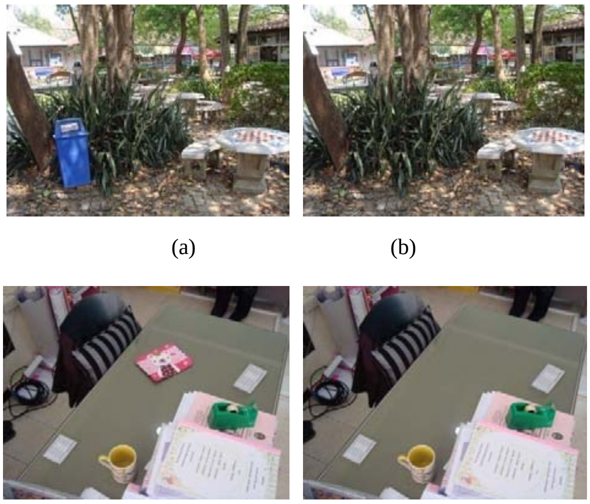
II. Proposed Method

Copy-Paste tampering is done by copying a region of the image and pasting it on another place in the same image. When a region is copied and pasted to another place, it will keep some of its underling features that can be used to identify tampering. The feature used here is the color pattern.

N. Chaitawittanun is with the Faculty of Information Science and Technology, Mahanakorn University of Technology, Bangkok 10530 Thailand (e-mail: nattapol_ctwtn@hotmail.com).

M. Munlin is with the Faculty of Information Science and Technology, Mahanakorn University of Technology, Bangkok 10530 Thailand (e-mail: mmunlin@gmail.com).

Specifically, we study the color present in an image where one of its regions is replicated, which has almost the same color pattern for both the copied and pasted parts. The general framework of our process is as shown in Fig. 5.



(a) (b)

(c) (d)

Fig. 3 Sample copy-paste image forgery (a) original image, (b) fake image, (c) original image, (d) fake image

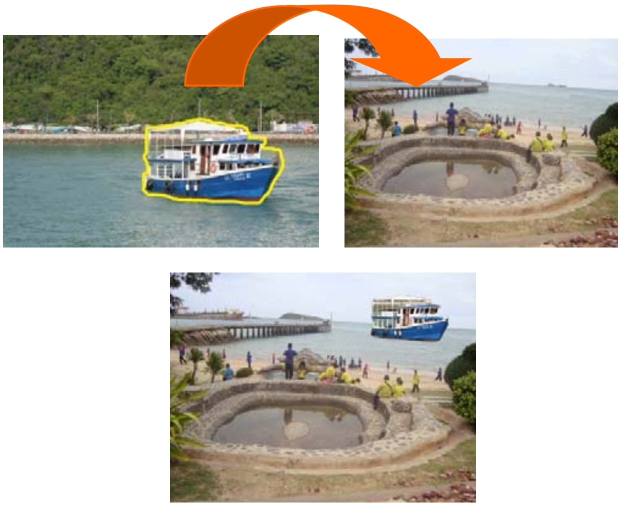


Fig. 4 Image splicing

A. Step 1: Decreasing Image Details

We first blur the image *f* for eliminate noise and detail, and then degrade the color of image.

B. Step 2: Splitting Image into Blocks NxN

Split the image *f*, of size m_f x n_f, which is tiled as blocks of pixels selected by sliding, pixel by pixel, from the top-left corner to the bottom-right corner.

C. Step 3: Extract Colors of Splitting Blocks and Cluster Data

We extract characteristic color with every block and categorize data of the image.

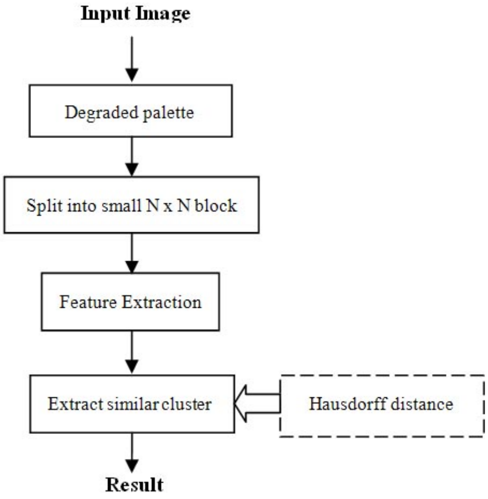


Fig. 5 The process of copy-move detection

D. Step4: Clustering the Similarity of Colors by Hausdorff Distance

Finding the similarity data is the process to identify a duplicate position by measuring distance of information group by Hausdorff distance. Given two finite point sets $A=\{a_1,\ldots,a_m\}$ and $B=\{b_1,\ldots,b_n\}$, the Hausdorff distance is defined as:

$$H(A, B) = \max (h(A,B),h(B,A)) \quad (1)$$

when

$$h(A,B) = \max_{a\in A} \min_{b\in B} \|a-b\| \quad (2)$$

and || || is some underlying norm on the points of *A* and *B*. The function *h (A, B)* is called the directed Hausdorff distance from *A* to *B*. It identifies the point a *A* that is farthest from any point of *B* and measures the distance from a to its nearest neighbor in *B*. The function *h (A, B)* in effect ranks each point of A based on its distance to the nearest point of *B* and then uses the largest ranked such point as the distance.

III. Experimental Result

We tested the performance of our proposed method on JPEG, TIFF, BMP, and PNG which have each of photo files are 100 images. All images are of 512x384 pixels. The photo was edited by copy-move technique as shown in Fig. 6.

Figs. 6 (a)-(c) show an example of copy-paste forgery. Fig. 6 (d) shows the modification of image is copied and pasted balloon in the image. Fig. 6 (e) shows weed picture is copied and pasted flower in the image. Fig. 6 (f) shows turtle picture is copied the upper turtle and pasted in the lower part.

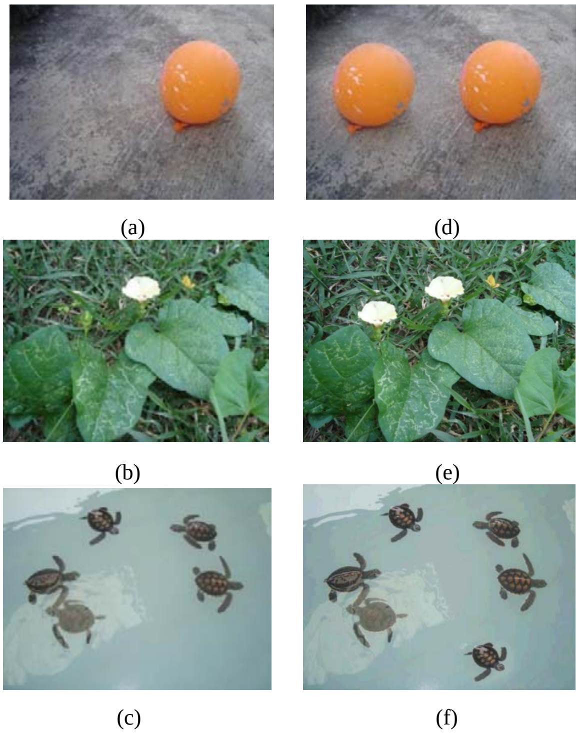

(a) (d)
(b) (e)
(c) (f)

Fig. 6 (a), (b), (c) original images (d), (e), (f) images tampered by copy-paste technique

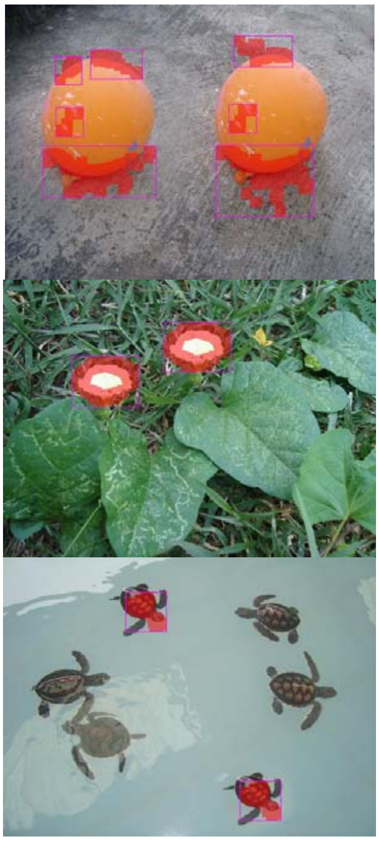


Fig. 7 Test images (Fig. 6 (d)-(f)) was the result of the detection

The results can be visually inspected in Fig. 7. The red areas depict the duplicated regions that were successfully detected. The measure of accuracy is summarized in Table I.

The result of the detection shows the accuracy of copy-paste technique of JPEG file which can be identified the duplicate position correctly at 64.36%. TIFF, BMP and PNG files show the accuracy at 62.54%, 61.78% and 60.98%. Average detection time of BMP file spent 120 seconds on a machine having Intel Core 2 Duo 2.1 GHz CPU and 4 GB RAM. In addition, JPEG, TIFF, and PNG spent 150, 180 and 165 seconds.

TABLE I
RESULTS OBTAINED FROM THE EXAMPLE FORGERIES

Image Type	% Accuracy Rate
JPEG	64.36
TIFF	62.54
BMP	60.98
PNG	61.78

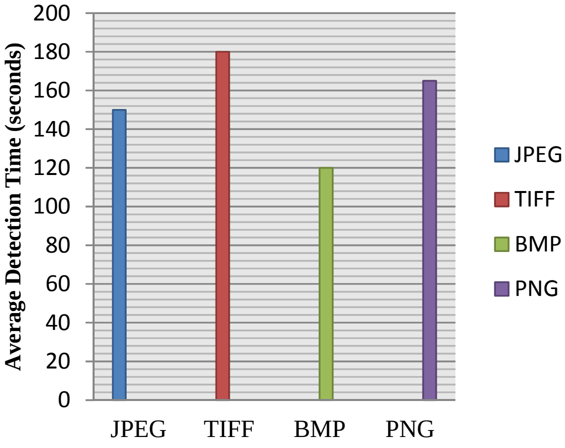


Fig. 8 Average detection time of each file format

IV. CONCLUSION

Digital image forgery detection is an interesting research topic in forensics science. An effective detection of specific copy-paste type of image tampering has been proposed in this paper. In this paper, we show that our process is useful to identify the copy-paste region. The proposed method can detect duplicated region from all sample images. In the future, we would like to detect other types of image files and enhance performance of the proposed detection.

REFERENCES

[1] L. Weiqi, Q. Zhenhua, P. Feng, H. Jiwu. "A survey of passive technology for digital image forensics". Springer Science. Frontiers of computer science in China.2007, vol. 1, no. 2, pp. 166-179.

[2] M.K. Johnsos. "Light and optical tools for image forensics". Ph.D. Thesis, Dartmouth College, Computer Science Deptl., Germany. 2007.

[3] D.A. Brugioni. "Photo fakery: the history and techniques of photographic deception and manipulation". Virginia :Brassey's publishers, 1999.

[4] Z. Lint, R. Wang, X. Tang, H.Y. Shum. "Detecting Doctored images using camera response normality and consistency". in Proc. Computer Vision and Pattern Recognition.2005, vol. 1, no. 43-48.

[5] B. Mahdian and S. Saic. "A bibliography on blind methods for identifying image forgery". ELSEVIER. Signal Processing: Image Communication.2010, vol. 25, Issue 6, pp. 389-399.

[6] I. Cox, M. Miller, J. Bloom. "Digital Watermarking : Principles & Practice". Morgan Kaufmann, 2001.

[7] J. Fridrich, D. Soukal, J. Lukas. "Detection of Copy-Move Forgery in Digital Images". in Proc. Digital Forensic Research Workshop, Cleveland, OH, Aug, 2003.

[8] A. Popescu and H. Farid. "Exposing digital forgeries in color filter array interpolated images". IEEE Trans. Signal Processing.2005, vol. 53, no.10, pp. 3948–3959.