Measurement of Copy-Move Forensic Image Similarity Using Distance Function

By ABDUL FADHIL

WORD COUNT 3017
TIME SUBMITTED 29-NOV-2019 02:51AM
PAPER ID 52514450

Measurement of Copy–Move Forensic Image Similarity Using Distance Function

Article *in* Journal of Computational and Theoretical Nanoscience · December 2018

DOI: 10.1166/asl.2018.12116

CITATIONS
0

READS
25

3 authors, including:

Abdul Fadlil
Ahmad Dahlan University
27 PUBLICATIONS **57** CITATIONS
SEE PROFILE

Imam Riadi
Ahmad Dahlan University
117 PUBLICATIONS **459** CITATIONS
SEE PROFILE

RESEARCH ARTICLE

Advanced Science Letters
Vol. 24, 9157–9162, 2018

Measurement of Copy-Move Forensic Image Similarity Using Distance Function

Abdul Fadlil[1,*], Imam Riadi[2], and Titi Sari[3]

[1]*Department of Electrical Engineering, University Ahmad Dahlan, Yogyakarta, Indonesia*
[2]*Department of Information System, University Ahmad Dahlan, Yogyakarta, Indonesia*
[3]*Magister of Informatics Engineering, University Ahmad Dahlan, Yogyakarta, Indonesia*

In digital imagery era, good-editing software allows users to process digital images in an easy way. As a result, image fraud practice is widespread inevitably. Therefore, an image-fraud-detection tool is required to prove the authenticity of a digital image. The growth of digital image manipulation has prompted writers in forensic imagery field to reveal the authenticity of images. In this research the original data images include two images object: beach.jpg and sailing.jpg. Then, manipulated images created manually: beach1.jpg and sailing1.jpg. The experiment is done through several stages of the conversion of color images into the form of grayscale degree followed by determining pattern through histogram of pixel image values. The image similarity is based on calculating distance value of two altered image patterns. Some of the distance functions performed in this research namely Euclidean distance, Manhattan distance and Canberra distance. While image manipulation used are limited into copy-move image manipulation. The Research results that distance function from two different images namely original image and manipulate image compare with original image and manipulated image's difference object. The experimental result of three distance functions show same result that smallest calculation value is distance calculation of original image and manipulated image from the same object.

Keywords: Image, Forensic, Copy-Move, Distance Function Method.

1. INTRODUCTION

The rapid growth of digital imaging technology has enabled high-resolution imaging devices at low cost. The presence of image processing software applications such as Corel Painter, Corel Paint Shop Pro, Corel Photo-paint, Adobe Photoshop, Microsoft Paint, and others make it easier for someone to manipulate images in accordance to his needs and desire. The ease of image processing using some of these software makes it possible to manipulate images that lead to crime. The emergence of such digital images may cause problems in the community's social life such as the loss of confidence in news, destruction of one's reputation, falsification of evidence and many other problems.[1] Some of the looming cases are including photos of the late North Korean President Kim Jong Il's funeral procession, floods in Korea, some officials checking broken roads in Huili China, photos of the 14th Street Bridge to commemorate the 30th anniversary of Air Florida Flight 90 in Washington Post, and a bird documentary pictures in the Sacramento Bee newspaper.[2]

Those cases prove the danger of image manipulation process for crime. This phenomenon should be our concern, as it will lead to harshness in the society. Because of the above cases, the public no longer trusts digital images and it demands the development of image forensic studies. Image forensics is a field of study that identifies the origin and verifies the authenticity of images. It can make a business field to trace crime in digital image world.

Image forensic research is categorized into two types: active and passive authentication. Active authentication requires additional information about the original image. This includes embedding process, watermarking into an image or extracting unique features as a mark of the image. Passive authentication is known as blind detection technique since it does not require any additional information about the original image. There are two categories of passive authentication. One that identifies the source device and other that detects image manipulation. Detecting image manipulation refers to the use of statistical analysis or techniques to detect forged areas.

2. LITERATURE REVIEW

An image definition is "a representation, likeness, or imitation of an object."[3] Image is grouped into visible and invisible images as referred in Figure 1.

Image is a function of light intensity of a two-dimensional object denoted in, $f(x, y)$ where x and y are coordinates of the

*Author to whom correspondence should be addressed.

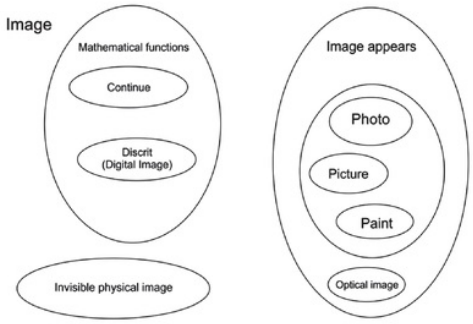

Fig. 1. Grouping of image types.

image point, while the value, $f(x, y)$ is the intensity level of the image at that point. Image function is expressed as follows:

$$I = f(yx) \tag{1}$$

Since $f(x, y)$ is function of light intensity, then $f(x, y)$ is a form of energy and thus has an intensity area from zero to infinity.

$$0 < f(x, y) < \infty \tag{2}$$

The advance of editing software makes it easy for a person to manipulate an original image without leaving its trace. Image manipulation can be categorized into three types; image splicing, copy-move image manipulation, and retouching images.[4] Below is the description:

Splicing image is the process of combining two or more images to create a new one. A particular region is copied from one image and inserted into another to form a different image.[5] The inequalities in the connected region can be directed to de-correlation detection. Examples of splicing images can be seen in Figure 2.

Copy-move image manipulation is a common type of image manipulation. It involves a process of copying and inserting in the same image.[6] The copied areas are generally modified using operations such as scaling, rotation, and adding to blend the manipulated areas with the surrounding. As a result, this manipulation is difficult to detect by human eye. An example of copy-move image manipulation can be seen in Figure 3.

Image retouching is a process of converting pixels that is copied according to the surrounding.[7] This can either improve or reduce some features of the original image without changing the actual meaning. The kind of manipulation is usually done by magazine editors to make an image more interesting.[8] Intrusion Detection System or abbreviated as IDS is a software application that can detect suspicious activity in a system or network. IDS can perform analysis and look for evidence of intrusion experiments (infiltration). Examples of retouching images can be seen in Figure 4.

Image manipulation cases are usually saved back and re-compressed as new JPEG images. Therefore, manipulation can be detected through the recompression. The periodic characteristic of JPEG images, both in spatial and domain altering, is suggested to be formulated in order to create a robust detection approach.[9] A-DJPG and NA-DJPG are two statistical models to illustrate the artifacts present in each type of recompression.[10]

Fig. 2. (a and b) original image, (c) splice image of (a and b).

3. METHODOLOGY

Image processing process in the research consists of image input process, image conversion, copy-move image manipulation, histrogram, and distance function process. Figure 5 describes the research flow.

The researchers use similarity measurement as the research methodology to measure the similarity of forensic images with distance function method. Similarity measurement is a process of measuring the resemblance of an object toward a reference object. In similarity measurement, distance measurement is performed, where higher the distance between two objects, the more different they are. Distance is usually the size of the unlike.[11] Euclidean Distance is an extension of the Phytagoras theorem on multidimensional data.[12] There are several popular distance functions and are often used in pattern recognition systems namely: Euclidean, Manhattan, Canberra, Bray-Curtis, Squared Chord, and Squared Chi-Squared.[13] In this study, the researchers limit on three distance functions: Euclidean, Manhattan, and Canberra.

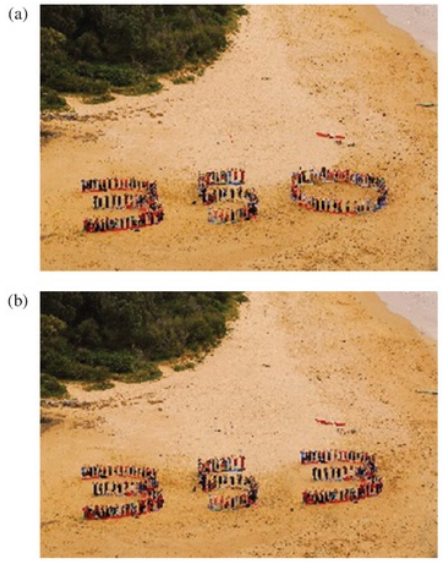

Fig. 3. (a) original image, (b) image copy-move manipulation of (a).

Euclidean distance is the sum of squares of two vector values (x, y)[14] and is defined as follows

$$d_E(x \cdot y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3)$$

Manhattan distance is the sum of the absolute function values of two vector values (x, y). It is also usually called as City-block distance. This method assumes that variables in cluster variant are not correlated.

$$d_M(x \cdot y) = \sum_{i=1}^{n}|x_i - y_i| \quad (4)$$

Fig. 4. (a) original image, (b) retouching image of (a).

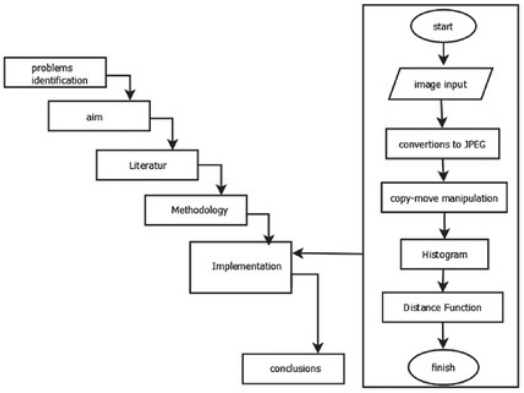

Fig. 5. Research flow chart.

Canberra distance is a frequently used metric function for data scattered around the origin. It was introduced in 1967.[15] Canberra metrics are similar to Manhattan spacing. The difference is that the absolute difference between variables of two objects is divided by the sum of the absolute variable values before the sum. The general equation is given in the form of:

$$d_C(x \cdot y) = \sum_{i=1}^{n}\frac{|x_i - y_i|}{|x_i| + |y_i|} \quad (5)$$

Using above calculation is expected to measure similarity of original image and copy-move manipulation image.

4. RESULTS AND DISCUSSION

The researcher will use a sample of original image obtained from Friedrich Alexander Universität Erlangen Nürnberg (FAU) website.[16] To get copy-move image manipulation, researchers deliberately manipulate original image by using adobe photoshopcs3 software application. By manipulated itself the original image will be emphasize which is completely manipulated from the original image.

Original image with name of beach.jpg manipulated by performing copy-move of part of region/object so that it has different impression than the original one. Result of manipulation is named beach1.jpg can be seen in Figure 6.

Similar to the image above. In this process, original image with name of sailing.jpg image is manipulated by conducting a copy-move on a particular sailboat. The result of copy-move image manipulation with name of sailing1.jpg can be seen in Figure 7.

Each image has a gray, red, blue and green color composition in each pixel. Color composition is something that will fill the color value in a pixel of a digital image. Any color changes on a single pixel will affect the value of image histogram.[17] Important information about the contents of a digital image can be determined by creating an image histogram. Histograms can also show

Fig. 6. (a) beach.jpg—original image, (b) beach1.jpg—image is manipulated copy-move.

a lot about the brightness and contrast of an image. An ideal encryption scheme image should produce a cipher image with a different histogram than a plain image.[18] A histogram can change an RGB image into grayscale one. Therefore, histogram is a valuable tool in image processing work, both qualitatively and quantitatively. Histogram code snippet is as follows:

```
g = imread('beach.jpg');
g1 = rgb2gray(g);
figure,imshow(g1);
figure,imhist(g1);
```

The display of grayscale histogram diagram for beach.jpg image is shown in Figure 8.

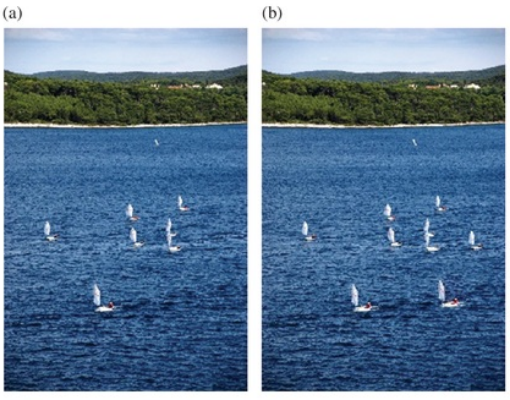

Fig. 7. (a) sailing.jpg—original image, (b) sailing1.jpg—image is manipulated copy-move.

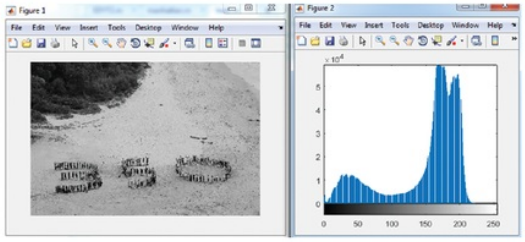

Fig. 8. beach.jpg histogram diagram.

The grayscale histogram diagram for A1.jpg copy-move image is shown in Figure 9.

The grayscale histogram diagram for sailing.jpg image is shown in Figure 10.

The grayscale histogram diagram for sailing1.jpg image is shown in Figure 11.

Based on the histogram data obtained, Euclidean Distance can be calculated to compare the similarity of two histograms of both images; the original and the copy-move images (beach.jpg and beach1.jpg). The Euclidean Distance code snippet is as follows:

```
Im1 = imread('beach.jpg');
Im2 = imread('beach1.jpg');
%plotting of th1emsubplot(1,2,1);
imshow(Im1);
subplot(1,2,2);
imshow(Im2);
Im1 = rgb2gray(Im1);
Im2 = rgb2gray(Im2);
%the code for conversion of image to its normalized histogram
x = imhist(Im1)./numel(Im1);
y = imhist(Im2)./numel(Im2);
% Calculate the Euclidean distance
E_distance = sqrt(sum((x - y)^2));
```

A Manhattan Distance calculation is performed to compare the similarity of the two histograms of both images; the original and the copy-move (beach.jpg and beach1.jpg).
The Manhattan Distance code snippet is as follows:

```
Im1 = imread('beach.jpg');
Im2 = imread('beach1.jpg');
%plotting of th1em
subplot(1,2,1);
imshow(Im1);
subplot(1,2,2);
imshow(Im2);
Im1 = rgb2gray(Im1);
```

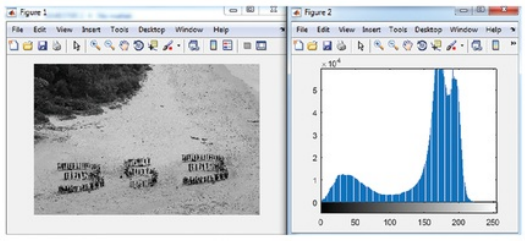

Fig. 9. beach1.jpg histogram diagram.

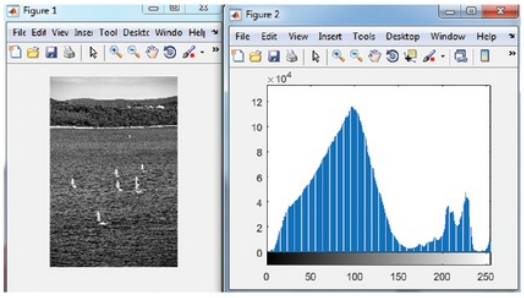

Fig. 10. sailing.jpg histogram diagram.

```
Im2 = rgb2gray(Im2);
%the code for conversion of image to its normalized histogram
x = imhist(Im1)./numel(Im1);
y = imhist(Im2)./numel(Im2);
% Calculate the Manhattan distance
M_distance = sum(abs(x - y));
```

Canberra Distance Calculation is conducted to compare the similarity of two histograms of both original and copy-move images (beach.jpg and beach1.jpg).

The Canberra Distance code snippet is as follows:

```
Im1 = imread('beach.jpg');
Im2 = imread('beach1.jpg');
%plotting of th1em
subplot(1,2,1);
imshow(Im1);
subplot(1,2,2);
imshow(Im2);
Im1 = rgb2gray(Im1);
Im2 = rgb2gray(Im2);
%the code for conversion of image to its normalized histogram
x = imhist(Im1)./numel(Im1);
y = imhist(Im2)./numel(Im2);
% Calculate the canberra distance
C_distance = sum(abs(x - y)./(abs(x) + abs(y)))
```

The Comparison of beach.jpg image and beach1.jpg image in Similarity Measurement calculation is as follows:

Table I it can be shown that distance value between beach.jpg and beach1.jpg is smaller in value than distance value between beach.jpg and sailing1.jpg.

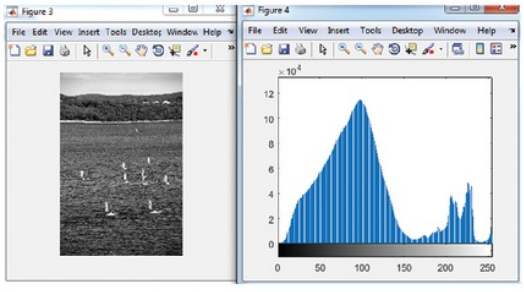

Fig. 11. sailing1.jpg histogram diagram.

Table I. Comparison of beach.jpg to beach1.jpg and sailing1.jpg.

	E_distance	M_distance	C_distance
	beach.jpg	beach.jpg	beach.jpg
beach1.jpg	0.0048	0.0434	21.6412
Sailing1.jpg	0.1148	1.3537	162.2907

Table II. Comparison of sailing.jpg to beach1.jpg and sailing1.jpg.

	E_distance	M_distance	C_distance
	sailing.jpg	sailing.jpg	sailing.jpg
beach1.jpg	0.1156	1.3578	163.0716
Sailing1.jpg	0.0013	0.0138	6.4432

Table II it can be shown that distance value between sailing.jpg and sailing1.jpg is smaller in value than distance value between sailing.jpg and beach1.jpg.

This research obtained the result that the size of similarity of an object can be known through calculation using distance functions. From three ways distance is: Euclidean Distance, Manhattan Distance, and Canberra Distance shows the same results with the calculation which is the visibility of the original image and the manipulated image.

5. CONCLUSIONS

The experiments that have been done can be seen that original image and manipulated copy-move images can show different histogram patterns. Similarity measurement can be done by calculating distance value between original image pattern and manipulated image compared with calculation of distance value between original image and image is manipulated with another object. Similarity measurement calculations show the same result that smallest calculated value is distance calculation of original image and manipulated image.

In subsequent research, experiments calculate the similarity of original image and image is manipulated with various manipulation variations of each object in hope that the method of distance functions will be implemented in measuring the similarity.

References and Notes

1. O. Al-Qershi and B. Khoo, Passive detection of copy-move forgery in digital images: State-of-the-art, Forensic Science International, Elsevier **(2013)**, pp. 284–295.
2. Fotografer.net, Accessed on May 2017.
3. W. T. Handoko, E. Ardhianto, and E. Safriliyanto, *Jurnal Teknologi Informasi DINAMIK* 16, 56 **(2011)**.
4. L. Weiyao, Qazi Tanzeela, Hayat Khizar, U. Khan Samee, A. Madani Sajjad, A. Khan Imran, Kołodziej Joanna, L. Hongxiang, Y. K. Choong, and Z. X. Cheng, *IET Image Processing* 7, 660 **(2013)**.
5. X. Zhao, J. Li, S. Li, and S. Wang, Detecting digital citra splicing in chroma spaces, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **(2011)**, Vol. 6526 LNCS, pp. 12–22.
6. Hussain Muhammad, Qasem Sahar, B. George, Muhammad Ghulam, Aboalsamh Hatim, and Mathkour Hassan, *International Journal on Artificial Intelligence Tools* XX, 1 **(2014)**.
7. D. R. Cok, "*Cloning Technique For Digital Citra Retouching*" **(1996)**.
8. Sadeghi Somayeh, A. Jalab Hamid, and Dadkhah Sajjad, Efficient copy-move forgery detection for digital citras, *World Acad. Sci. Eng. Technol.*, World Academy of Science, Engineering and Technology **(2012)**, Vol. 71, pp. 542–546.
9. Y.-L Chen and C.-T. Hsu, *IEEE Transactions on Information Forensics and Security* 6.

10. B. Tiziona and P. Alexandro, *IEEE Signal Processing Society* 7 **(2012)**.
11. A. C. Rancher, Methods of Multivariate Analysis, Second edn., John Wiley & Sons, Canada **(2004)**.
12. Hair Jr., F. Joseph, Black, C. William, Babin, C. Barry, and R. E. Anderson, Multivariate Data Analysis 7/e, Pearson Prentice Hall, New Jersey **(2010)**.
13. Fadlil Abdul, Sistem Pengenalan Citra Jenis–Jenis Textile, Spektrum Industri **(2012)**, Vol. 10, pp. 1–107.
14. Yampolskiy, P. Roman, and Venu Govindaraju, *International Journal of Biological and Life Sciences* 1, 227 **(2005)**.
15. G. N. Lance and W. T. Williams, *Australian Computer Journal* 1, 15 **(1967)**.
16. https://www5.cs.fau.de/research/data/image-manipulation/, Accessed on May 2017.
17. W. Hendro, R. Imam, and P. Yudi, Encryption EXIF Metadata for Protection Photographic Image of Copyright Piracy, IJRCCT **(2016)**, Vol. 5.
18. Irfan Pahrul, Prayudi Yudi, and Riadi Imam, *International Journal of Computer Applications (0975-8887)* 123 **(2015)**.

Received: 19 June 2017. Accepted: 22 July 2017.

Measurement of Copy-Move Forensic Image Similarity Using Distance Function

ORIGINALITY REPORT

0%

SIMILARITY INDEX

PRIMARY SOURCES

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		