Image Forensic for detecting Splicing Image with Distance Function

Imam Riadi
Department of Information System, Ahmad Dahlan University, Yogyakarta

Abdul Fadlil
Department of Electric Engineering, Ahmad Dahlan University, Yogyakarta

Titi Sari
Magister of Information System, Ahmad Dahlan University, Yogyakarta

ABSTRACT
In the era of digital image, good editing software allows users to process digital images in an easy way. It is inevitable, which, unfortunately leads to the widespread of image forgery. Hence, an image fraud detection tool is essential to verify the authenticity of a digital image. The rapid growth of digital image manipulation has prompted writers on forensic image to reveal their authenticity. Manipulations are commonly found in image formats such as Joint Photographic Experts Group (JPEG). JPEG is the most common format supported by devices and apps. Therefore, the researchers will analyze measurement of forensic image similarity using distance function method, while image manipulation is used specially on image splicing. The results of this study show that distance function can be 2 different images.

General Terms
Digital Forensics

Keywords
Image, Forensics, Splicing, Distance Function.

1. INTRODUCTION
The rapid growth of digital imaging technology has enabled high-resolution imaging devices at low cost. Image processing application software 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 as needed.

The ease of image processing with some of the software makes it possible to manipulate images that lead to crime. For example, there is a case where a young artist's photographs were manipulated into vulgar ones (showbiz.liputan6.com, July 31, 2015). Similar case also happened to a member of Kepri archipelago's House of Representative (tribunbatam.com, 18 August 2015) and President Jokowi when he met Anak Dalam tribe (bali.tribunnews.com, 3 November 2015).

The examples above proof us the danger of image manipulation process for crime. Hence, we should put this as our concern as it will lead to harshness in the society. As a result of the above cases, public may no longer trust digital images and it triggers the emergence of image forensic studies. Image forensics is a study that identifies the origin and verifies the authenticity of an image. Image forensics can make a business field to trace crime in digital image world.

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

This research will implement forensic image resemblance measuring method with distance function method. The research is focused on image manipulation that is image splicing and two types of distance functions that are Euclidean and Manhattan distance. The image files used are JPEG files.

The research aims to compare manipulated images that are most resemble to the original image through a forensic process. Based on the forensic analysis conducted, image evidence can be used as one of the auxiliary elements in a digital image crime disclosure.

The scope of discussion of the research is image manipulation, particularly image splicing and image resemblance measurement with distance function using Matlab to create forensic-analysis-related .toolbox.

2. LITERATURE REVIEW
2.1 Digital Image
The definition of image is "a representation, likeness, or imitation of an object" [1]. it can be grouped into visible and invisible image.

Images can be grouped into visible and invisible images as referred in Figure 1.

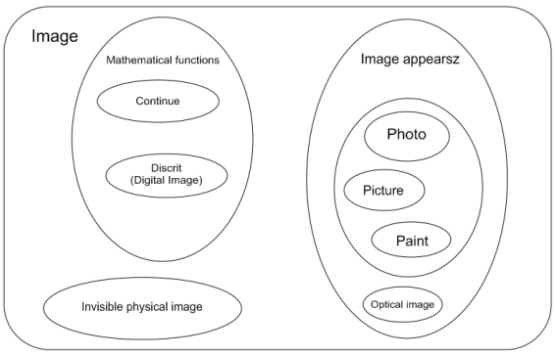


Fig 1. Grouping of image types

Image is a function of light intensity of a two-dimensional object denoted in , $f(x, y)$ where x and y are the image point's coordinates, whereas, $f(x, y)$ is the intensity level of the image at that point. The image function is expressed as follows:

$$I = f(y\ x) \qquad (1)$$

Since *f (x, y)* is a function of light intensity, then *f (x, y)* is a form of energy, so that it has an area of intensity from zero to infinity.

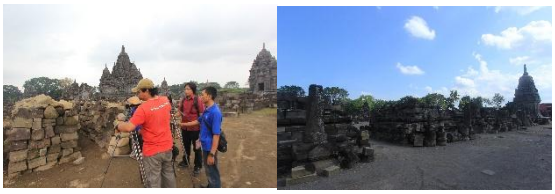
$$0 < f(x,y) < \infty \qquad (2)$$

2.2 Forensics Images

The advancement of editing software makes it easier for a person to manipulate an original image without leaving any trace. Image manipulation can be categorized into three types; Image splicing, copy-move image manipulation, and retouching images [2], as described below:

2.2.1 Splicing Image

Splicing image is a process of combining two or more images to create a new one. A particular area is copied from one image and inserted into another to form a different one [3]. The inequalities in the connected region can be directed to de-correlation detection. Examples of splicing images can be seen in the following figure 2:



(a) (b)

(c)

Fig 2. (a) A1.jpg original image, (b) A2.jpg original image, and (c) A splice.jpg forgery splicing of both A1.jpg and A2.jpg images

2.2.2 Copy-move Image Manipulation

Copy-move is a common type of image manipulation. It involves a copied and inserted process in the same image [4]. The copied areas are generally modified by operations such as scaling, rotation, and adding to blend the manipulated area with the surrounding area. As a result, this manipulation is difficult to detect by the human eye. An example of *copy-move* image manipulation can be seen in the figure 3:

(a) (b)

Fig 3. (a) A1.jpg original image, (b) A copy. jpg forgery copy-move image from A1.jpg

2.2.3 Retouching Image

Retouching Image is a process of converting pixels which are copied according to the surrounding ones [5]. This can either improve or reduce some features of the original image without changing the actual meaning. This type of manipulation is usually done by magazine editors to make an image more interesting [6]. Intrusion Detection System or abbreviated as IDS is a software application that can detect suspicious activities in a system or network. IDS can perform an analysis and search for evidence of any intrusion experiments (infiltration). Examples of retouching images can be seen in figure 4:

(a) (b)

Fig 4. (a) A1.jpg original image, (b) A retouch.jpg forgery retouching image from A1.jpg.

2.3 JPEG Compression in Image Forensics

Image manipulation cases are usually re-saved and re-compressed as new JPEG images. Therefore, manipulation can be detected through the recompression. Error rate will increase on re-save operation. After the next re-save operation, square grid may reach minimum level of error[7]. The periodic characteristics of JPEG images both in spatial and domain alterations are suggested to formulate in order to create a robust detection approach [8]. Statistical models are to illustrate artefacts both of A-DJPG as well as NA-DJPG for each type of recompression [9].

3. METHODOLOGY

Research methodology used by researchers to measure the similarity of forensic images with the method of distance function using Similarity Measurement. Similarity Measurement is the process of measuring the resemblance of an object to a reference object. In Similarity Measurement, distance measurement will be performed. The more significant the distance between two objects, the more distinct they are. Distance is generally the size of unlike [10].

Euclidean Distance is basically an extension of the Phytagoras Theorem on multidimensional data [11]. *Euclidean distance* is the sum of squares of two vector values (x, y) [12] and is defined as follows

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

Manhattan distance is the sum of an absolute value function of two vector values (*x, y*). It is also commonly called as City-block distance [13]. This method assumes that variables in variate cluster are not correlated.

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

4. EXPERIMENTAL RESULTS

4.1 Citra Splice

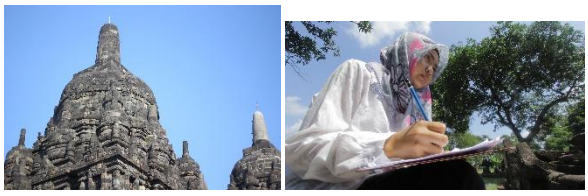
A1.JPG and A2.JPG images are manipulated by combining human or people and surrounding objects. Then, it is combined with a second image that is a blue sky object. It results into a new image of A splice.JPG image as seen in Figure 5.

(a) (b)

(c)

Fig 5. (a) A1.jpg original image, (b) A2.jpg original image, and (c) A splice.jpg forgery splicing both of A1.jpg and A2.jpg images

Similar to the above image, on the following process C2.JPG image is combined C1.JPG image where the person's object is taken and then combined into image splicing manipulation, C Splice.JPG image as in Figure 6.



(a) (b)

(c)

Fig 6. (a) C1.jpg original image, (b) C2.jpg original image, and (c) C splice.jpg forgery splicing of both C1.jpg and C2.jpg images

4.2 Image Processing

In this research, image processing process which consists of splice image input, image conversion, image splicing, histrogram, and similarity measurement processes will be performed using Matlab 2016b software. Diagram of research flow as in Figure 7.

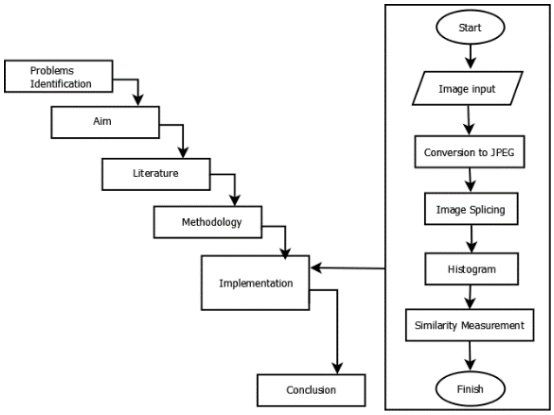


Fig 7. Diagram of Research Flow

Important information about the contents of a digital image can be revealed by creating an image histogram. Histograms can also show a lot of information about the brightness and contrast of an image. Histogram may change RGB image into grayscale one. Therefore, histogram is a valuable tool in image processing work both qualitatively and quantitatively.

The histogram code snippet is as follows;

```
g = imread('A1.jpg');
g1=rgb2gray(g);
figure,imshow(g1);
figure,imhist(g1);
```

The grayscale histogram diagram display for A1.jpg image is shown in Figure 8.

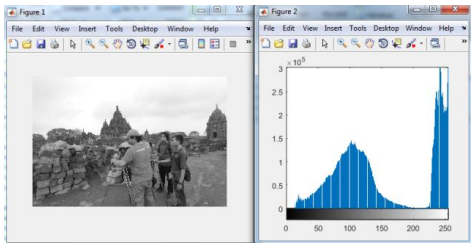


Fig 8. Histogram diagram of A1.jpg

The grayscale histogram diagram display for the A2.jpg image is as shown in Figure 9.

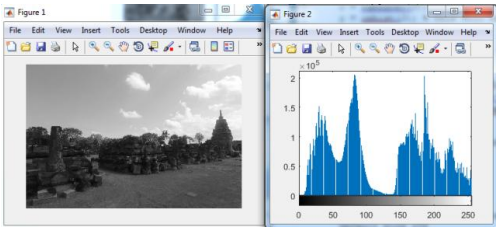


Fig 9. Histogram diagram of A2.jpg

Below is the grayscale histogram diagram of A Splice.jpg image as shown in Figure 10.

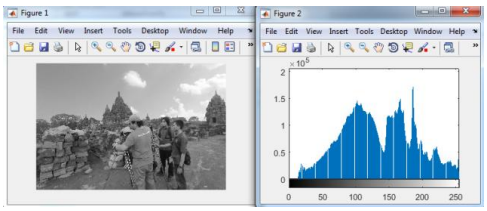


Fig 10. Histogram diagram of A Splice.jpg

Below is the grayscale histogram diagram of C1.jpg image as shown in Figure 11.

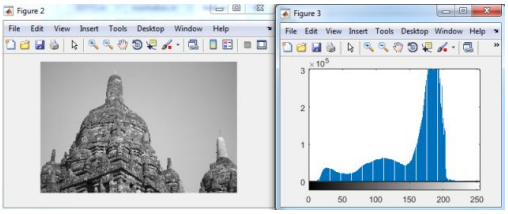


Fig 11. C1.jpg Histogram diagram

The following diagram shows the grayscale histogram for C2.jpg image as shown in Figure 12.

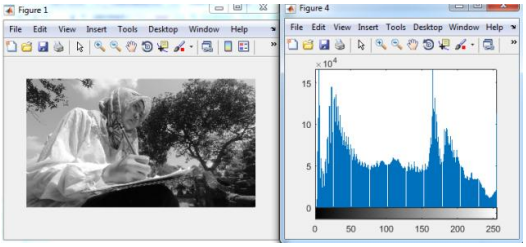


Fig 12. C2.jpg histogram diagram

The following diagram shows the grayscale histogram for CSplice.jpg image as shown in Figure 13.

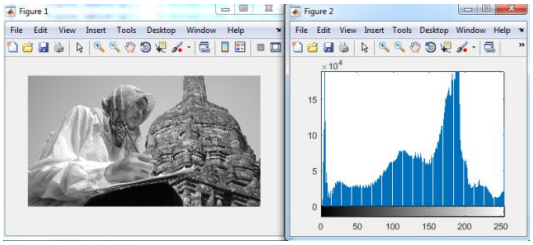


Fig 13. C Splice.jpg histogram diagram

Once the histogram data is obtained, the researcher can calculate Euclidean Distance to compare the similarity of the two histograms of both images that is splice image and the original image (A1.jpg and A2.jpg). Below is the snippet of Euclidean Distance code:

```
Im1 = imread('A splice .jpg');
Im2 = imread('A1.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 Euclidean distance
E_distance = sqrt(sum((x-y).^2));
```

The Manhattan Distance's calculation is used to compare the resemblance of both two holograms that is, splicing image and the original (A1.jpg and A2.jpg). Below is the snippet of Manhattan Distance code:

```
Im1 = imread('A splice .jpg');
Im2 = imread('A1.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 Manhattan distance
M_distance = sum(abs(x-y));
```

The comparison results from A Splice.jpg with A1.jpg, and A2.jpg images with Similarity Measurement calculations can be seen in table 1

Table1. Comparison Similarity Measurement Result A Splice of A1.jpg and A2.jpg

	E_distance	M_distance
	A Splice.jpg	A Splice.jpg
A1.jpg	0.0956	0.6980
A2.jpg	0.0531	0.6394

The comparison results of C Splice.jpg with C1.jpg and C2.jpg images with Similarity Measurements calculation can be seen in table 2

Table 2. Comparison Similarity Measurement Result C Splice of C1,jpg and C2.jpg

	E_distance	M_distance
	C Splice.jpg	C Splice.jpg
C1.jpg	0.0529	0.4939
C2.jpg	0.0463	0.4879

5. CONCLUSION

In the experiments conducted, the methods used by the researchers starting from combining images/splice image, processing images in histogram data and calculating are Euclidean distance and Manhattan distance.

Finally, it can be concluded that the results of image similarity measurement shows that A Splice.jpg is resemblance to A2.jpg image while C Splice.jpg resembles C2.jpg image. It is revealed by smaller Euclidean and Manhattan Distance values.

From the sample data of splice images that are compared to its original, researchers found ascending distance-based similarity measurement calculation results both on Euclidean and Manhattan distance methods.

6. REFERENCES

[1] Handoko W.T., Ardhianto E. dan E. Safriliyanto. 2011. *"Analisis Dan Imp lementasi Image Denoising dengan Metode Normal Shrink sebagai Wavelet Thresholding Analysis"*. Jurnal Teknologi Informasi DINAMIK. 16 (1): 56 - 63.

[2] Lin Weiyao, Qazi Tanzeela, Hayat Khizar, Khan Samee U., Madani Sajjad A., Khan Imran A., Kołodziej Joanna, Li Hongxiang, Yow Kin Choong, Zhong Xu Cheng, 2013 "*Survey on blind image forgery detection*,", Volume 7, Issue 7, 2013 , p. 660 – 670, IET Image Processing.

[3] Zhao X., J. Li, S. Li, and S. Wang, 2011*," Detecting digital citra splicing in chroma spaces, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)"*, vol. 6526 LNCS, pp. 12–22.

[4] Hussain Muhammad, Qasem Sahar, bebis George, Muhammad Ghulam, Aboalsamh Hatim, Mathkour Hassan, 2014, "*Evaluation of image forgery detection using multi-scale Weber local descriptors"*Vol. XX, No. X (2014) 1–27, International Journal on Artificial Intelligence Tools

[5] Cok D. R., 1996, "*Cloning Technique For Digital Citra Retouching*,".

[6] Sadeghi Somayeh, Jalab Hamid A., and Dadkhah Sajjad, 2012 "*Efficient Copy-Move Forgery Detection for Digital Citras*," World Acad. Sci. Eng. Technol., vol. 71, no. 11, pp. 542–546, World Academy of Science, Engineering and Technology

[7] Sari Titi, Riadi Imam, Fadlil Imam, 2016. *"Image Forensics for File Engineering Detection Using Error Level Analysis"*, ISBN : 979-587-626-0 UNSRI Vol-2, No.1(2016), Annual Research Seminar (ARS)

[8] Chen. Y.-L and Hsu. C.-T, 2-11 "*Detecting Recompression of JPEG Images via Periodicity Analysis of Compression Artifacts for Tampering Detection*", Volume: 6, Issue: 2. IEEE Transactions on Information Forensics and Security.

[9] Bianchi Tiziona and Piva Alexandro, 2012. , "*Image Forgery Localization via Block-Grained Analysis of JPEG Artifacts"*,Volume: 7, Issue: 3, IEEE Signal Processing Society

[10] Rancher, A . C . 2004. *"Methods of Multivariate Analysis Second Edition"* . Joh Wiley& Sons, Canada.

[11] Hair Jr., Joseph F., Black, William C., Babin, Barry C., dan Ro lph E. Anderson. 2010. *"Multivariate Data Analysis 7/e".* Pearson Prentice Hall, New Jersey.

[12] Yampolskiy, Roman P. and Venu Govindaraju. 2005. *"Similarity Measure Functions for Strategy Based Bio metrics".* International Journal of Biological and Life Sciences . 1(4): 227- 228.

[13] Royce Winston, 1970, *"Managing The Development of Large Software Systems"*, page 1-9, IEEE Wascon.