Who Recognized Fake Image Better: A Comparison Study between Human and Computer

R Bahana¹, A Nurhudatiana², M Hidajat¹ and T A Budiono¹

¹Computer Science Program, Bina Nusantara University, Jakarta, Indonesia
²Information Systems Program, Bina Nusantara University, Jakarta, Indonesia

E-mail: rbahana@binus.edu

Abstract. Along with the rapid development of multimedia and information technology in the recent years, a large number of photos and videos can be easily created and distributed by each and every member of our society. Some of these images are genuine, while some others have gone through modifications. Although the creation and posting of edited photographs in media are basically not prohibited, it becomes a problem when they are used improperly. Until now, the common practice in Indonesia when assessing whether an image has been modified or is original is still heavily dependent on human expert assessment that could be subjective. This research offers an alternative solution that is objective by developing a systematic image analysis method for distinguishing between original and modified image. The method utilizes high-level feature, namely the metadata of the image file, to differentiate between original and modified image. A prototype system was implemented in Visual Basic as a desktop application. The application used MS Windows GDI+. This prototype application can be used as a reliable and faster means to assist in fake image detection. However, as metadata-based analysis only detects whether a photograph has possibly been edited and it does not show which part of the photograph has been manipulated, future work includes development of pixel-based analysis approach.

1. Introduction
Multimedia technology nowadays has made it very easy for anyone to create and share edited images publicly on the Internet. While many of these images are created with no ill intention (e.g., humor), there are fake images intentionally produced to create rumors, sensations, or even defamation of public figures. Fig. 1 (a) and (b) show examples of easily recognized fake images with humor intention (e.g., celebrity meme) [1] while Fig. 1(c) shows another fake image widely circulated on the Internet known to be a form of character attack to Abraham Samad, the previous Chairman of the Corruption Eradication Commission [2]. Unlike the typical meme images in Fig. 1(a) and (b) which seem unrealistic, Fig. 1(c) looks more natural and can be deceiving to people who see it, especially those without any background in photography or image/video analysis.

When controversial photographs become viral in public, it is often not easy to recognize whether the image is actually real or hoax [3]. The common practice in Indonesia to handle such issue is to ask human experts to give their analyses on the photographs. The problem with this approach is that this kind of subjective assessment may lead to varying conclusions depending on each expert’s background and experience. Furthermore, with the massive growth of manipulated images on the Internet, the speed of visual evaluation by human eyes will not catch up with the speed of the production of the images themselves. This research aims to provide a preliminary study on how a
simple fake image detection system can be used to assist in the recognition of fake images in the public media. In this research, we hypothesized that the automated system would at least give a comparable performance to human assessment in terms of accuracy with significantly faster performance.

![Example Images](image_url)

**Figure 1.** Examples of manipulated images obtained from Google Search (Pinterest and blogs)

### 2. A Metadata-Based Fake Image Detection System

#### 2.1. Metadata-based approach

Many image forgery detection algorithms have been proposed in literature [4-11]. In general, these techniques can be categorized into high-level feature analysis, where analysis is performed on the whole image as a file (e.g., file metadata), and low-level feature analysis, where analysis is based on parts of the file (e.g., image pixels). Some of the features proposed for use in pixel-based techniques include the use of motion blur metric [9], color filter array analysis [10], and direction of sunlight [11].

Metadata contains information on the structure and properties of a file. The information contained within the EXIF (Exchangeable Image File Format) metadata of a photograph typically contains information as follows: camera models, the manufacturer, and the settings at the time the picture is taken, such as aperture, shutter speed, focal length, metering mode, and ISO speed, the characteristics of the image, the date and time the picture taken, and a thumbnail or a mini version of the image [12]. An example of publicly available image forensic service which uses metadata to check input image is Ghiro [13]. Metadata also contains information about the file compression rate. Since a lot of useful information is available in metadata, it can be used to identify the source file and processing history that ever happened to the file. According to the study by Kee et al. [14], JPEG Headers can provide useful information for metadata-based analysis. JPEG is a compressed format and it is the most common format for storing image files. Every camera model basically produces JPEG Headers information differently from one another [14].

#### 2.2. Prototype of the system

The system was implemented as a desktop application by using Visual Basic programming language. The application used MS Windows GDI+, which is a class-based API. The Microsoft Windows graphics device interface (GDI) enables applications to use graphics and formatted text on both the video display and the printer. Applications based on the Microsoft Win32 API do not access graphics hardware directly. Instead, GDI+ interacts with device drivers on behalf of applications. GDI+ is also supported by Microsoft Win64. GDI+ stores an individual piece of metadata in a PropertyItems object. The PropertyItems is used to retrieve metadata information from a file in the format of array. EXIF information is one of the information contained in PropertyItem Objects. A flowchart to illustrate the method is given in Fig. 2.

For each image, the system takes approximately 1 second to produce a binary output: 0 (original) or 1 (possibly manipulated). Through the Graphical User Interface (GUI), user only needs to open an image or a photo by clicking the button “Open Image” and choose an image or photo he/she wants to
check. When the image is original, it will show output like in Fig. 3(a). If the image no longer contains EXIF information, which means that the EXIF metadata has been lost, the image might have been altered or engineered, and thus has possibly been modified (see Fig. 3(b)).

3. Results and Analysis

For this study, Computer Science students were recruited as our human experts. These students were 3rd year students who had taken Multimedia course, so they already had basic knowledge of image editing. Two stages of manual detections were conducted. In the first stage, 15 pictures (i.e., 6 modified and 9 original) were given to 21 panelists. The panelists were asked to determine: 1) whether the image was original or manipulated, and 2) if they thought that the image had been manipulated, which part of the image seemed to have been modified. The difficulty level across the dataset was varied.

![Flowchart](image)

**Figure 2.** Flowchart of the modified image detection method
In the second stage of the experiment, 10 pairs of pictures were shown side by side and 6 panelists were asked to determine which side the engineered image was and to indicate the element of the image which seemed to have been modified. The image datasets used for this research are available in the following two URLs [15, 16].

Table 1 shows the percentage of correct and incorrect answers made by the panelists for each image in Dataset 1. From the 15 images displayed to the panelists, most of them (71% of them) incorrectly guessed photo number 6 as original (see Fig. 4). The license plate of the motorcycle in photo number 6 had actually been changed by duplicating the first character from “B” to “BB”. For the image with the most correct answer, 95% of the panelists answered correctly for image number 14 because the trace of copy-move was very visible. In average, only 61% of test participants answered correctly for Dataset 1. This shows that it can be very difficult to recognize manipulated images.

Table 2 shows the percentage of correct and incorrect answers made by the panellists for each image in Dataset 2. From the 10 pairs of images displayed as test materials, most of the test participants (87%) answered correctly for image number 4 (see Fig. 5(a) for the original image and Fig. 5(b) for the modified image) and more (60%) gave incorrect answers for image number 7 (Fig. 6(a) for the original image and Fig. 6(b) for the modified image). In average, 67% of test participants answered correctly, slightly higher compared to the first stage of experiment. A possible reason is because in the second stage, the panellist viewed the two photos altogether, so they could compare both photos and they already knew that one was original and another one was manipulated. The result shows that even by showing the original photo as reference, humans could still find it difficult to recognize edited images.

**Table 1.** Percentage of correct and incorrect answers – Test I

| Photo | % correct answer | % wrong answer |
|-------|-----------------|---------------|
| 1     | 71%             | 29%           |
| 2     | 43%             | 57%           |
| 3     | 33%             | 67%           |
| 4     | 62%             | 38%           |
| 5     | 67%             | 33%           |
| 6     | 29%             | 71%           |
| 7     | 71%             | 29%           |
| 8     | 57%             | 43%           |
| 9     | 52%             | 48%           |
| 10    | 38%             | 62%           |
| 11    | 90%             | 10%           |
| 12    | 71%             | 29%           |
| 13    | 57%             | 43%           |
| 14    | 95%             | 5%            |
| 15    | 71%             | 29%           |
| Average | 61%           | 39%           |

**Table 2.** Percentage of correct and incorrect answers – Test II

| Photo | % correct answer | % wrong answer |
|-------|-----------------|---------------|
| 1     | 67%             | 33%           |
| 2     | 73%             | 27%           |
| 3     | 67%             | 33%           |
| 4     | 87%             | 13%           |
| 5     | 60%             | 40%           |
| 6     | 73%             | 27%           |
| 7     | 40%             | 60%           |
| 8     | 47%             | 53%           |
| 9     | 80%             | 20%           |
| 10    | 73%             | 27%           |
| Average | 67%           | 33%           |
Figure 4. Photo number 6 in Dataset 1 is the most difficult to recognize

Figure 5. (a) original photo of image number 4 in Dataset 2; (b) a green tile was copied and moved to the blue tiles

Figure 6. (a) original photo of image number 7 in Dataset 2; (b) blue light was duplicated into 3 lines from originally 1 line

Based on the correct answer from Table 2, when the panelists were asked to point which part in the image had been edited, the results are shown in Table 3. It can be seen that only 59% could point out the edited location in the image correctly. There is a possibility that the panelists only guessed the answers during second stage as they could not point out the right location where the image had been edited. With the sophistication of image-editing applications which can manipulate photos easily and are accessible by many people, this could be a worrying thing, especially when the images are intended for a crime or spreading wrong perceptions in society. Educating the general public would not be easy without the help of an application that can detect the originality of an image.
Table 3. Percentage of correct and wrong place – Test II

| Photo | % correct place | % wrong place |
|-------|-----------------|---------------|
| 1     | 87%             | 13%           |
| 2     | 73%             | 27%           |
| 3     | 80%             | 20%           |
| 4     | 87%             | 13%           |
| 5     | 67%             | 33%           |
| 6     | 27%             | 73%           |
| 7     | 73%             | 27%           |
| 8     | 67%             | 33%           |
| 9     | 73%             | 27%           |
| Average | 59%           | 41%           |

3.1. Comparison with the prototype
All photos in the first and second stages of experiment had already been checked using the prototype application. The prototype application could detect the EXIF information from all original photos and did not display any EXIF metadata information when edited photos were inputted.

4. Conclusion and Recommendation
In this research, we have developed an alternative solution to the current practice of manual recognition of manipulated images by developing a simple Visual Basic system which can be used as a reliable and faster means to assist in fake image detection. One flaw of this method is that metadata is so easy to access, making it easy to remove or replace. In addition, metadata-based analysis can only suspect whether a photograph has possibly been edited and it does not show which part of the photograph has been manipulated. Future direction of this research would be to implement content-based approach to give more fine-grained analysis of the photograph.

Acknowledgment
This research is supported by DIKTI Hibah Bersaing research funding 2016.

References
[1] Kaskus 2014 Super-ngakak-meme-foto-syahrini-tiduran-di-padang-rumput-di-itali-yang-unyu-unyu Retrieved from http://www.kaskus.co.id/thread/53ce6cf45a5163b46f8b46f8/super-ngakak-meme-foto-syahrini-tiduran-di-padang-rumput-di-itali-yang-unyu-unyu/82
[2] Global Indonesian Voices 2015 Retrieved from http://www.globalindonesianvoices.com/18576/foto-mesra-ketua-kpk-abraham-samad-says-intimate-photos-with-elvira-devinamira-are-fake/
[3] Forester T and Morrison P 1994 Computer ethics: Cautionary tales and ethical dilemmas in computing, 2nd ed Cambridge MA: MIT Press
[4] Scientific Working Group Imaging Technology 2007 Section 14: Best practices for image authentication Retrieved from http://www.crime-scene-investigator.net/swgit-section14.pdf
[5] Farid H 2009 A survey of image forgery detection IEEE Signal Processing Magazine, 2(26), 16-25
[6] Sencar H and Memon N 2009 Overview of state-of-the-art in digital image forensics Algorithms, Architectures and Information Systems Security, p 325
[7] Mahdian B and Saic S 2010 A bibliography on blind methods for identifying image forgery. Image Communication, 25(6), 389-399
[8] Mahdian B, Saic S and Nedbal R 2010 JPEG quantization table forensics: A statistical approach. Proceedings of International Workshop on Computational Forensics, 150-159
[9] Kakar P, Sudha N, and Ser W 2011 Exposing digital image forgeries by detecting discrepancies in motion blur IEEE Transactions on Multimedia, 13(3), 443-452
[10] Popescu A C and Farid H 2005 Exposing digital forgeries in color filter array interpolated images IEEE Transactions on Image Processing, 53(10), 3948-3959
[11] Kakar P and Sudha N 2012 Authenticating image metadata elements using geolocation information and sun direction estimation. Proceedings of International Conference on Multimedia and Expo (ICME), 236-241
[12] Metadata Working Group 2010 Guidelines for handling image metadata v2.0. Retrieved from http://www.metadataworkinggroup.org/specs/
[13] Ghiro [Software]. Available from http://imageforensic.org/
[14] Kee E, Johnson M K and Farid H 2011 Digital image authentication from JPEG headers IEEE Transactions on Information Forensics and Security, 6(3), 1066-1075
[15] Dataset 1. Available: http://raymondbahana.blog.binusian.org/rekayasa-photo-i
[16] Dataset 2. Available: http://raymondbahana.blog.binusian.org/rekayasa-photo-ii