Effectiveness of Probabilistic Image Sampling Techniques to Identify Hoax-related Images in Indonesia

C. W. D. Lumoindong, M. A. Aryadi, I. T. Wilyani, A. Suhartomo

Abstract*: Hoaxes are very common among Indonesians. The tendency of most Indonesians to believe everything they saw or heard and the rapid spread of information with questionable credibility in social media contribute to the quick growth of hoaxes. These hoaxes varies from the 'light' hoaxes such as April Fools pranks which are taken seriously to some 'heavier' ones such as political hoaxes. Fortunately, there are a lot of websites offering hoax identification services. But, most Indonesians would rather holding on the term 'no picture means hoax' than checking any kinds of information they received on those websites. As image editing software progressed forward, this old term is not really helpful. Forged images are easily made and spread through social media, and only few Indonesians know how to distinguish between real images and forged images. This research will focus on comparing the probabilistic image sampling techniques in order to combat hoaxes spreading through social media. Before being identified, several images (both forged and real) alongside some opinion-based questions regarding hoax-related imagery will be presented in a form of a survey to 167 respondents, in which most of respondents failed to identify the forged images. The success of the probabilistic image sampling technique will be based on the detection test score of each sampling techniques and their suitability with current situation in Indonesia.

***Keywords* : Hoax, Probabilistic; Image processing; Forgery**

I. INTRODUCTION

In the past five decades the development of digital image processing has grown significantly. Since the computer as the main device to do the digital image processing is developing very fast, its process can be handled easily. That means, almost everyone can forge images to their own likings as long as they have the required tools.

Many people tend to benefit themselves by forging images, especially in a legal document such as ID card and some legal certificates [1]. Indirectly, the development of computer technology also encourage the spread of image forgery. New technologies enable people to easily forge documents with user-friendly software, such as Adobe Photoshop.

Revised Manuscript Received on January 5, 2020.
***** Correspondence Author
C. W. D. Lumoindong*, Electrical Engineering Department, President University, Cikarang, Indonesia. Email: dlumoindong@ymail.com
M. A. Aryadi, Electrical Engineering Department, President University, Cikarang, Indonesia. Email: ariefaryadi383@gmail.com
I. T. Wilyani, Electrical Engineering Department, President University, Cikarang, Indonesia. Email: indahtriawilyani@yahoo.com
A. Suhartomo, Electrical Engineering Department, President University, Cikarang, Indonesia. Email: asuharto@president.ac.id

Masyarakat Telematika Indonesia or Indonesian Telematics Community (Mastel) conducted a survey about hoaxes in 2017 [2]. This research main objective is to have an image of what are Indonesian people interpretation to hoax, the classification and also the effect of the hoax itself.

According to the survey, Indonesians agreed that hoax is a fake news made by purpose and most of them know it after they see the clarification. Indonesians who know about the true fact behind the hoax itself is only 14.4% [2]. When Indonesians received a message which they perceived as 'news', most of them will check the truth behind the news by consulting to a search engine such as Google. They will spread the 'news' easily if the senders are one their relatives or people that they trusted.

Social media made a major impact to the spreading of hoax in Indonesia. Social media helped create a 'collective reality' in which the voice of the majority always treated as reality [3]. This 'collective reality' also supported by the fact that most Indonesians are very active on social media. In Indonesia, social media plays an important role in both higher welfare and higher cybercrime rate, including image forgeries [4].

The skyrocketing growth of hoaxes and the increase of their resemblance to the true ones depends on one major factor, which is image. The hoax that came from an image contributed 37.5% of all hoaxes [2]. Old hoaxes usually comes with very few to no images. The quality of the provided images are also abysmal due to the limitations of image-editing software. But the recent innovations in digital imaging software have proven beneficial to the growth of hoaxes [5]. With those new innovations, one will find that the differences of a forged and a real image are very hard to point.

Actually, photojournalists are allowed to edit their photos as long as it follows the ethical editing rules. These rules limit the permitted editing to color, brightness and composition adjustments (such as cropping), while tampering and/or omitting a part of an image is strictly prohibited [6]. Most of the forged images are clearly violate these rules, as they are usually tampered, cloned, and/or partially omitted to the forgers' liking.

From the late sixties to the early seventies many researchers did researches to prove the originality of an image with several methods. The methods used in those

researches vary from a simple crosscheck to more complex way such as utilizing computer vision. However, there has not been a research which successfully found a one-stop, fool-proof way to identify fake images.

This research is aimed to find out whether probabilistic image sampling techniques, such as NOI4, ELA, and SIFT are good enough to provide a reliable forgery detection service. The parameter of success will be based on the detection test score, with a score of 150 and above is seen as effective. The survey test is also used to determine whether these sampling techniques are suitable to be used to counter growing hoaxes in Indonesia.

II. THEORETICAL BACKGROUND

A. Types of Image Forgeries

Despite their likeness to an original image, forged images still have some considerable difference compared to the original true image. One of them is "ghosting", or a pattern of incoherent pixels, in the edited areas [7]. If the forger didn't pay attention to the details, forged images can be identified by looking at the traces in the image. Another sign of forged image is the non-uniform noise patterns, or commonly known as "grains". But, an image of high resolution is needed in order to precisely determine the originality of an image by its grains, since compression will render the grains less visible.

Forged images are usually found in form of montage. If the images used in the montage have different resolutions, the image can be identified by revealing each zones' "compression histories" [8]. The common method to check the compression histories is by comparing the target image to several compressed versions of itself. If the target image is a montage, then there will be a constant difference of resolution between the montage elements.

The second type is the realistic additive forgery, in which an image is forged by adding an element taken from a similar image. This similar image can be taken on the same day and/or same location with same resolution. This type of forgery is harder to detect, since it looks almost like unforged original image.

The third is the copy-move forgery, or known as cloning. Cloning can be done in many ways, and sometimes the results may vary from the realistic-looking one to a very poorly executed forgeries. This type of forgery has been covered by a lot of researchers [9] [10].

B. Forged Image Detection Methods

In terms of detection methods, one of the most widely used is by checking the SIFT (Scale-Invariant Feature Transform) keypoints of the image. SIFT keypoints utilizes frequency sampling of the source images with 1% added noise so that a random number will be added to each pixel [11]. Keypoints are descripted by identifying the gradient of each point in the image.

The keypoints generated by SIFT method are matched using a process based on nearest-neighbor matching to determine image originality. This matching process follows the Bayesian Theorem to recognize objects [12]. Bayesian theorem is proven effective in recognizing objects, even in high-level computer vision application [13].

However, SIFT requires a pre-defined image description as keypoint comparator. So, in order to check the originality of an image, one should find (or at least, speculate) the 'original' image to serve as keypoints source. Then, the generated keypoints from the 'original' image are compared to the suspected 'forged' image.

The other method is by Error Level Analysis (ELA). This method basically compresses an image by a specific margin of error, utilizing JPEG loss in its process [14]. After several compressions, the compressed images will be compared to each other. If the image was forged, then it will show constant differences in the forged regions. From the research conducted by Krawetz, ELA is proven to be effective in identifying images forged from multiple images with different resolutions [14]. For example, it is shown in Fig. 1 that the image on the right (analyzed image) implies that the words are copy-moved to the kid's shirt (see the white area on the analyzed image).

Fig. 1. Example of ELA identification (Krawetz, 2007).

On the other hand, Zampoglou et al stated that NOI4 is a superior algorithm compared to ELA [15]. This is based on conducted tests involving several datasets utilizing the same MATLAB code used by Pan et al in 2014 [15]. NOI4 is described as a method to detect forged image by checking the Gaussian noise pattern of different regions, or usually known as median-filter noise residue inconsistencies [15]. NOI4 relies on the variance of the native noise of the image to differ it to the added noise. The equation used to estimate the variance of noise is shown in (1).

$$\hat{k}_x, \hat{\sigma}_n^2 = \arg\min_{k_x, \sigma_n^2} \sum_{i=2}^{N^2} \left| \frac{K_x - 3}{\left(1 + \frac{\sigma_n^2}{\sigma_{yi}^2 - \sigma_n^2}\right)} + 3 - K_{yi} \right|^2 \quad (1)$$

With Kx as the kurtosis of image x, σ_n^2 as the noise

variance, and *yi* as the response image.

III. METHODS

This research used the SIFT-forensic MATLAB code [8], NOI4 image forensics MATLAB code [15], and *Forensically* online application.

Forensically is used as a comparator due to its usage of adjustable ELA algorithm. The results of all methods are compared and analyzed according to the complexity and the accuracy of identification.

The images used in this research are obtained from personal archive (Fig. 2(a) and 2(c) to provide control over the results, while for SIFT test, the SIFT image profile used is from Amerini et al [9]. The image in Fig. 2(b) is slightly-altered test image from the same research. The images are shown in Fig. 2.

(a)

(b)

(c)

Fig. 2. Images used in this research.

Image *(a)* is a forged image with cloned boat and added raft. The raft added to image *(a)* is from another image taken on a similar place. This image represents the realistic additive forging technique, where two image taken on a similar place composed a montage.

Image *(b)* is a test image from the SIFT copy-move experiment [9], which was modified by cloning the bookshelves, walls, and books on the floor. The bookshelves are copy-moved to simulate realistic additive technique, the walls are cloned (with *Adobe Photoshop*'s clone stamp tool) and the books on the floor are copied and resized to simulate montaging technique

Image *(c)* represents the montaging image forgery technique. The original image is an image of a street painter and his collection taken on Braga Street, Bandung. The image was later forged by adding The Starry Night painting by Van Gogh and The Scream by Edvard Munch. All the forged areas of the test images are red-lined in Fig. 3.

(a)

(b)

(c)

Fig. 3. Forged areas of the images.

Before being analyzed, the images are shown to 167 respondents in form of a survey. The survey also asked about their news preferences and the impact of imagery to article reliability. The survey form used to collect the data is shown in Fig. 4 and divided into 3 opinion-based questions and 5 hoax test questions, in which 5 images are provided with 3 of them being forged images.

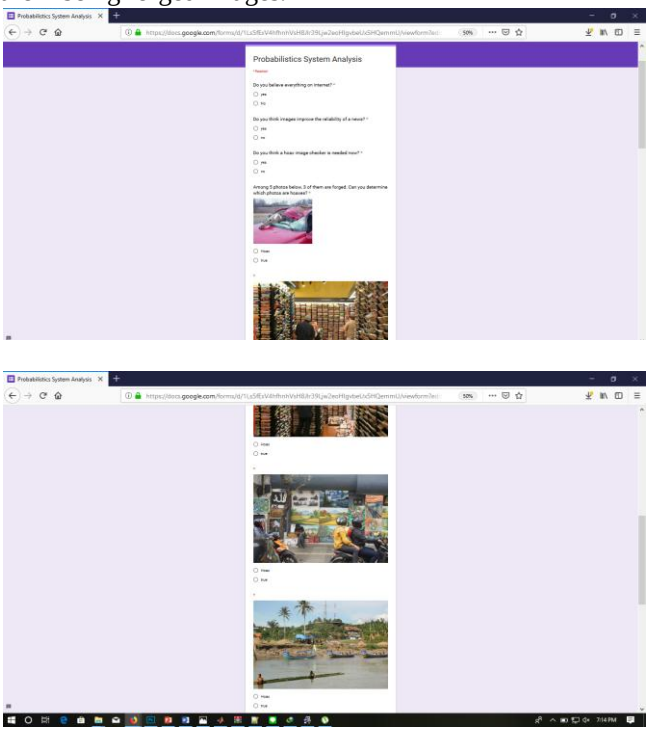


Fig. 4. The form used in this research.

IV. RESULT AND DISCUSSIONS

The result of analysis process by both MATLAB and *Forensically* is shown in Table-I. *PD* stands for *Partially Detected*, *D* for *Detected*, and *ND* for *Not Detected*. Note that the SIFT test can only be done for image (*b)* due to its nature as a comparative test.

Fig. 5. Table- I: Detection Test Results

Image	ELA	NOI4	SIFT[a]
(a)	*PD*	*PD*	-
(b)	*D*	*ND*	*D*
(c)	*PD*	*D*	-

[a] SIFT only tested on image (*b*).

In order to measure the effectiveness of each methods, a numerical test score data of Table 1 is generated. *PD* is expressed as 50, *D* is expressed as 100, and *ND* is expressed as 0. The numerical data is shown in Table-II.

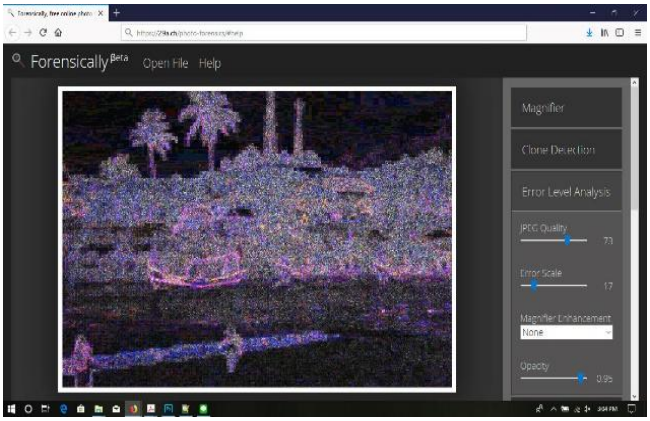
Table- II: Detection Test Scores

Image	ELA	NOI4	SIFT[a]
(a)	50	50	-
(b)	100	0	100
(c)	50	100	-
TOTAL	200	150	100

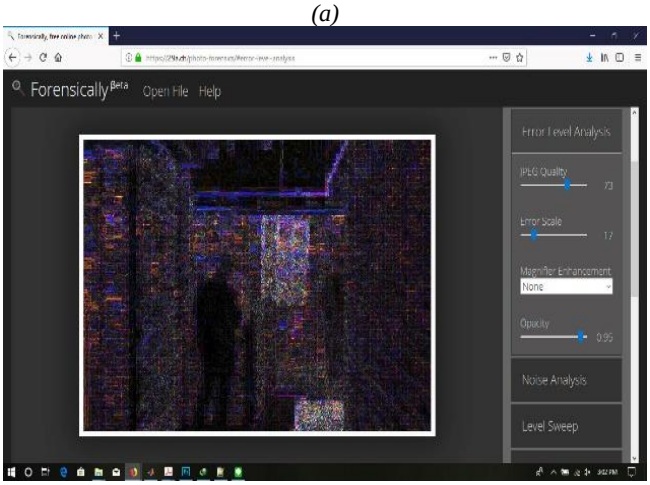
[a] SIFT only tested on image (*b*).

The Forensically test results and their corresponding ELA configurations can be seen in Fig. 5. Note that the forged areas detected by this method will have different brightness compared to its neighboring parts.

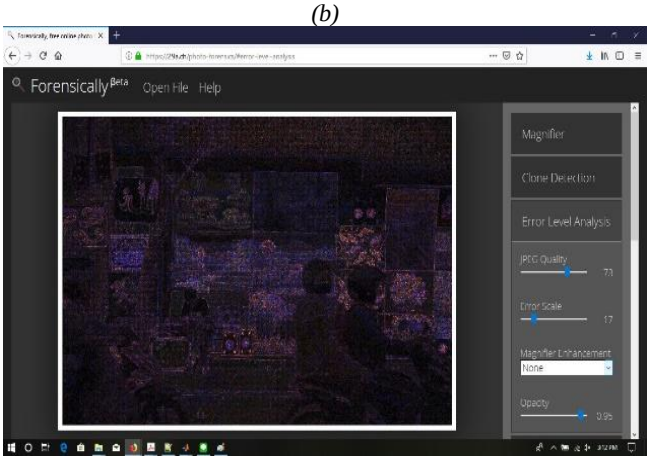
From the results, ELA did well in detecting forgeries in image *(b)* and *(c)*, thanks to some resizing and skewing which are applied on both images forged components to fit well. These transformations (resizing, skewing) tends to lower the quality of the image (or in this case, part of image). However, image *(a)* is also forged using part of itself (boat part), which means the overall quality of the forged part is almost the same with the base image. Thus, ELA struggles to detect it, although successfully isolate part of the raft part.



(a)



(b)

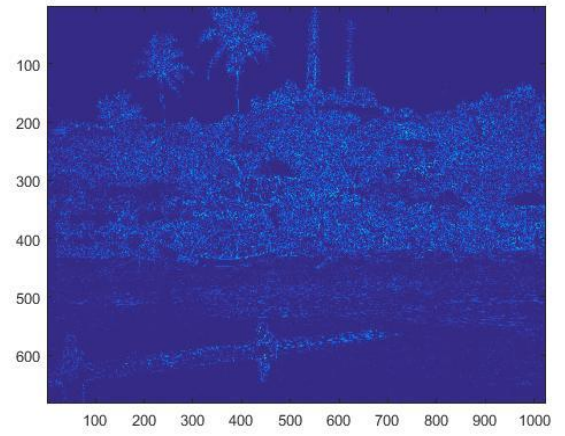


(c)

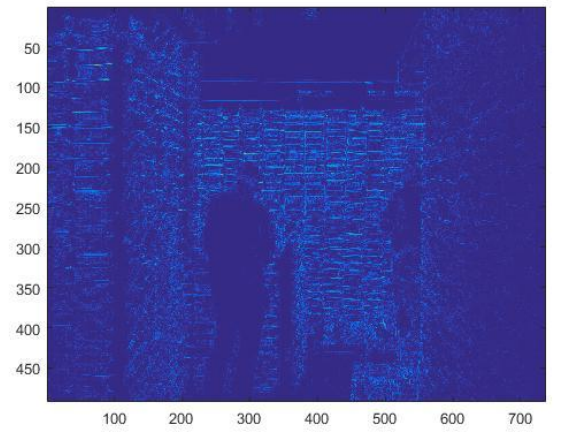
Fig. 6. Forensically ELA identification results.

The MATLAB generated

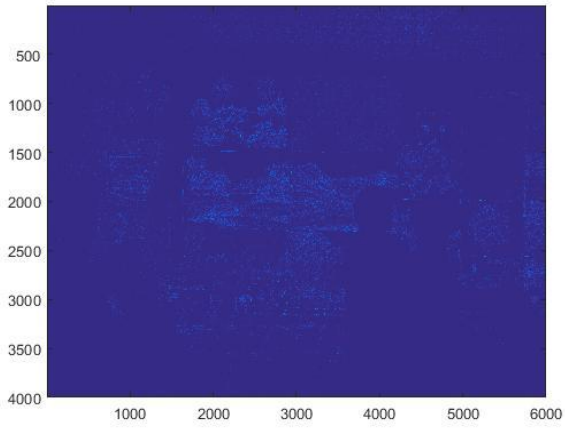
figure of NOI4 detection results is shown in Fig. 6. Note that in Fig. 6 *(a)*, the raft has a visibly different noise pattern compared to the rest of the images, and so with the two paintings in Fig. 6 *(c)*, which are displayed as plain blue blocks.



(a)



(b)



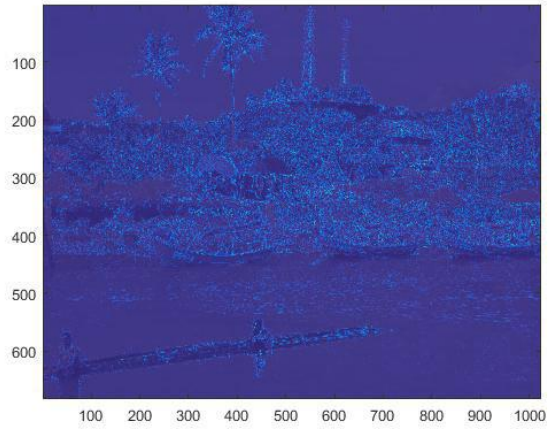
(c)

Fig. 7. NOI4 detection results.

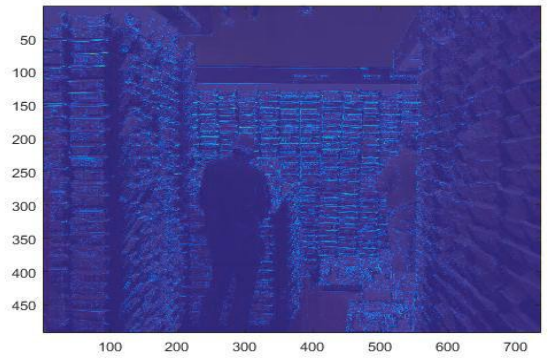
To highlight the forged areas, their corresponding images will be added as overlays. The detection results with images overlay is shown in Fig. 7.

From the results, NOI4 works best if the forged part is cloned or added from other image (such as image *(a)* and *(c)*), so that the forged part will show a different noise pattern. This is why Zampoglou et al stated that the NOI4 is the best-performing method in their research, since their research was focused on image tampering by adding parts of other image (montaging) [15]. However, Image *(b)* highlights one of NOI4 weakness. NOI4 did not perform well on detecting forgery in the said image since all of the forgeries performed on image *(b)* are from its own parts, thus making all noise patterns inside it look like a natural image.

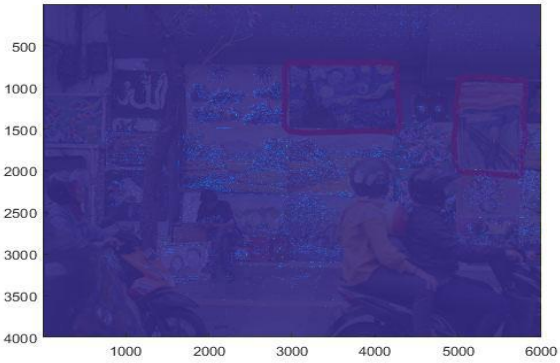
The SIFT result is shown in Fig. 8. The SIFT result showed some possible duplication of objects in a photo, thanks to the use of Bayesian object recognition. However, the duplication of books on the lower end and the duplication of wall on the upper end are not detected.



(a)



(b)



(c)

Fig. 8. NOI4 detection results with 10% image overlay.

Fig. 9. SIFT detection result.

The survey results of the questions and hoax tests are shown in Table-III and Table-IV, respectively.

Table-III. Opinion-based questions result.

Question	Yes (%)	No (%)
Do you believe in everything on the internet?	21	79
Do you agree that images improve the reliability of a news?	72	28
Do you think a hoax image checker is needed now?	85	15

Fig. 10. Table-IV. Opinion-based questions result.

Image	Hoax	True	Actual	Majority Guess
1	78.4%	21.6%	True	Wrong
2	15.6%	84.4%	Hoax	Wrong
3	12%	88%	Hoax	Wrong
4	39.5%	60.5%	Hoax	Wrong
5	61.1%	38.9%	True	Wrong

Based on the results of the detection tests, all the methods have their own strength and weaknesses. NOI4 will perform well if the components of the image have different noise pattern. ELA works best against components with different qualities. SIFT has a potential to be an all-rounded performer, but since this method relies heavily on matching parts of an image to a predefined description, it is not practical to be used as a sole method to detect image forgery.

However, this research is based on the performance of NOI4, ELA, and SIFT on only three different forging types (realistic additive, montaging, and cloning) and only done to three images (one images with three techniques in case of SIFT). So it cannot be seen as a total accurate representation of each methods effectiveness. For example, NOI4 method has a lower total detection test score (150) compared to ELA (200), but perform better on image *(c)*, in which NOI4 got 100 and ELA got 50. Also, the programs used to do the NOI4 and SIFT detections are experimental codes, which means it may contain some bugs or limitations.

Despite all their strength and weaknesses, each method have proven to be one step ahead compared to normal human vision. This is based on the hoax test results, where the majority of respondents failed to identify which images are forged. Test score-wise, both ELA and NOI4 achieve scores over 150, so both of them are effective to provide reliable image forgery detection services. As for the SIFT, even with 100 score, it still considered effective since it successfully detected the forgery in image *(b)*.

V. CONCLUSION AND FUTURE WORKS

All methods (NOI4, ELA, and SIFT) are considered as viable options to effectively detect image forgeries. However, they must be combined to achieve the best results, at least until a better image forgery detection method is invented.

Despite all their weaknesses, implementation of NOI4, ELA, and SIFT in one combined system of image forgery detection is a promising option to combat rapidly growing hoaxes in Indonesia. This is due to the majority of respondents agreed that images can increase the reliability of a material (in this case, news). If a hoax image detector service can exist as demanded by the majority of respondents, then hoaxes can be suppressed and national stability will also increase.

ACKNOWLEDGMENT

The researchers would like to thank all the authors whose works are cited in this paper and all the respondents of the survey for making this research happens.

REFERENCES

1. Kaur, A., Saran, V., & Gupta, A. K. Digital image processing for forensic analysis of fabricated documents, International Journal of Advanced Research in Science, Engineering and Technology, 2014 1, 2, 83-89.
2. Masyarakat Telematika Indonesia, Tanpa hoax Indonesia sejahtera. Available: https://www.bkkbn.go.id/po-content/uploads/Infografis_Hasil_Survey_MASTEL_tentang_Wabah_Hoax_Nasional.pdf (in Indonesian)
3. Situngkir, H., Spread of hoax in social media, BFI Working Paper No. WP-4-2011, 2011
4. Setiawan, T. & Suhartomo, A. The relation between internet use and societal development in Indonesia, International Conference on Sustainable Engineering and Creative Computing (ICSECC), 2019 133-137
5. Sarma, B., & Nandi, G. A Study on digital image forgery detection. International Journal, 2014, 4, 11.
6. Agence France-Presse. AFP editorial standards and best practices. [Online] Available: https://www.afp.com/sites/default/files/paragraphrich/201701/22_june_2016_afp_ethic.pdf
7. Redi, J. A., Taktak, W., & Dugelay, J. Digital forensics: a booklet for beginners. Multimedia Tools and Applications, 2011, 51, 133–162
8. Farid, H. Exposing digital forgeries from jpeg ghosts, IEEE Transactions on Information Forensics and Security. 2019 [Online] Available: http://www.ists.dartmouth.edu/library/434.pdf
9. Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A., & Serra, G. A SIFT-based forensic method for copy-move attack detection and transformation recovery, IEEE Transactions on Information Forensics and Security. 2011, 6, 3, 1099-1110.
10. Warbhe, A. D., Dharaskar, R. V., & Thakare, V. M. A scale invariant digital image copy-paste forgery detection approach based on NCC, International Journal of Computer Applications. International Conference on Communication, Computing, and Virtualization. 2015
11. Lowe, D. G. Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision. 2004, 60, 2, 91-110.

12. Chang, L., Duarte, M. M., Sucar, L. E., & Morales, E. F. Object class recognition using SIFT and Bayesian networks, Mexican International Conference on Artificial Intelligence (MICAI) 2010: Advances in Soft Computing. 2010, 56-66.
13. Pirri, F. & Romano, M. A Situation-Bayes View of Object Recognition Based on SymGeons. 2003, [Online] Available: https://www.researchgate.net/profile/Fiora_Pirri2/publication/2893651_A_Situation-Bayes_View_of_Object_Recognition_Based_on_SymGeons/links/00b4952ca86e93e82a000000/A-Situation-Bayes-View-of-Object-Recognition-Based-on-SymGeons.pdf?origin=publication_detail
14. Krawetz, N. A Picture's Worth: Digital Image Analysis and Forensics. 2007 [Online] Available: https://www.hackerfactor.com/papers/bh-usa-07-krawetz-wp.pdf
15. Zampoglou, M., Papadopoulos, S., & Kompatsiaris, Y. Large-scale evaluation of splicing localization algorithms for web images, Multimedia Tools and Applications. 2017, 76, 4, 4801-4834.

AUTHORS PROFILE

C. W. D. Lumoindong is a final-year undergraduate student in the Department of Electrical Engineering, President University, Cikarang, Indonesia. His area of research interests include Robotics, Image processing, and Wireless communication. He acquired the basic knowledge of Image processing based on his self-interest. He has published 2 research papers and presented 1 paper in national conference.

M. A. Aryadi, is a final-year undergraduate student in the Department of Electrical Engineering, President University, Cikarang, Indonesia. His area of research interests include Instrumentation, Image processing, and Mobile communication.

I. T. Wilyani, is a final-year undergraduate student in the Department of Electrical Engineering, President University, Cikarang, Indonesia. His area of research interests include Instrumentation, Image processing, and Mobile communication.

Drs. A. Suhartomo, is currently working as the Head of Electrical Engineering Study Program in the Faculty of Engineering, President University, Cikarang, Indonesia. He completed his Bachelor's degree in Physics in Universitas Sumatera Utara (North Sumatera University), Medan, Indonesia, his Master's degree in Electrical Engineering in the University of Indonesia, Jakarta, Indonesia and his Ph. D degree in Electrical Engineering in Stevens Institute of Technology, New Jersey, USA. His area of research include Fiber optic telecommunication, Network design, Automation, and Entrepreneurship. He had 27 years of industrial experience and 12 years of lecturing experience. He is an active member of both IEEE (Institute of Electrical and Electronics Engineer) and IAENG (International Association of Engineers).