

Kajian Penerapan Semi-Automated Evaluation Based on Similarity pada Investigasi Digital Image Forensics

Astrid Lestari Tungadi*, Erick Alfons Lisangan**

Universitas Atma Jaya Makassar

*astrid_tungadi@lecturer.uajm.ac.id, **erick_lisangan@lecturer.uajm.ac.id

ABSTRACT

Along with the development of information technology, is now widely available for image processing applications are easy to use. It has an impact on the process of manipulation and modification of the image to be more easily so it is quite difficult to determine the authenticity of an image. Image evaluation methods have been developed automatically. The weaknesses are irrelevant parts of the image will be processed so as not efficient in terms of processing time. User Role (semi-automatic) for assessing the relevance of an image section can be helpful in the evaluation method. Similarity-based approach, one of them using the Euclidean distance, either automatic or semi-automatic allows forensic computer analysis process becomes faster. Preparation of the ranking of the data that has similarities allows users to investigate the authenticity of the picture. In this paper tried to assess the implementation of semi-automatic evaluation of similarity based on a forensic image. In the method of analysis, there are two stages, namely the conduct of the evaluation results ranking image block that requires verification of the user so that the second stage is only relevant image blocks are analyzed automatically. The results showed that the use of semi-automatic evaluation methods based on computer forensic similarity not only be done on the data string but also the image data.

Keyword: *pixel-based technique, semi-automated, similarity, digital image forensics*

1. Introduction

Saat ini tidak ada penetapan metodologi untuk memverifikasi keaslian dan integritas citra digital secara otomatis. Mendeteksi pemalsuan pada citra digital merupakan bidang penelitian yang penting untuk memastikan kredibilitas gambar digital [1]. Hal tersebut disebabkan karena banyaknya citra visual yang dapat dilihat pada majalah untuk industri fashion, jurnal ilmiah, kampanye politik, ruang sidang, dan foto *hoax* yang masuk ke dalam *inbox* email, hingga foto kedokteran yang muncul dengan frekuensi dan kecanggihan yang semakin pesat. Di sisi lain, integritas dan tingkat kepercayaan dari citra visual tersebut tidak dapat dipercaya dan dibuktikan secara langsung seiring perkembangan teknologi digital [2]. Banyaknya aplikasi pemrosesan citra yang *user friendly* memungkinkan proses manipulasi dan modifikasi data citra mudah dilakukan dan sulit untuk mengetahui keasliannya dengan menggunakan mata manusia [3][1]

Teknik deteksi pemalsuan citra digital diklasifikasikan menjadi pendekatan aktif dan pasif. Pada pendekatan aktif, citra digital memerlukan beberapa pra-pengolahan seperti *watermark embedding* pada saat membuat gambar. Pendekatan aktif tidak dapat digunakan untuk melakukan otentifikasi citra digital. Pendekatan pasif digunakan atas asumsi bahwa pemalsuan citra digital mungkin tidak meninggalkan petunjuk visual tetapi mengubah statistika yang terdapat pada citra [2]. Salah satu teknik deteksi pemalsuan citra digital yang diperkenalkan oleh Farid (2009) adalah *pixel-based technique* dimana teknik tersebut mendeteksi anomali statistika pada tingkat *pixel*.

Saat ini telah banyak teknik deteksi pemalsuan citra digital secara otomatis. Salah satunya teknik yang diperkenalkan oleh Deshpande dan Kanikar (2012), yaitu *Fast Copy-Move Forgery Detection*. Pada teknik berbasis pada *pixel* tersebut dikhususkan untuk mendeteksi bagian dari gambar yang di-*copy* dan disisipkan (*move*) pada gambar itu sendiri [1]. Apabila hanya bagian kecil dari gambar yang mengalami proses *copy-move* maka berdampak pada waktu pemrosesan yang tidak efisien karena beberapa bagian seharusnya tidak perlu diinvestigasi tetap diproses.

Hankins dan Liu (2014) memperkenalkan suatu metode evaluasi berbasis similaritas yang sifatnya semi otomatis dengan tetap membutuhkan *feedback* dari pengguna. *Feedback* yang dibutuhkan dengan melakukan verifikasi terhadap bagian yang telah memiliki nilai similaritas apakah relevan dengan dokumen yang diperbandingkan. Apabila suatu bagian tidak relevan menurut pengguna maka bagian tersebut tidak

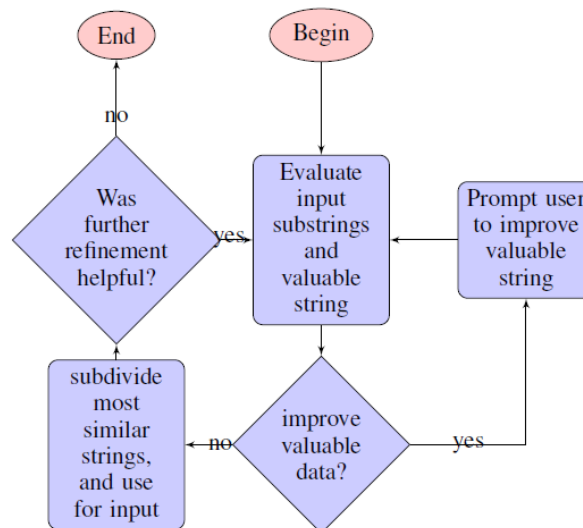
akan diproses pada tahap selanjutnya. Pemberian *feedback* dari pengguna memberikan keuntungan dimana bagian yang diproses merupakan bagian yang benar-benar relevan terhadap data yang dibandingkan [4].

Pada penelitian ini akan dikaji penerapan evaluasi semi otomatis berbasis similaritas yang diperkenalkan oleh Hankins dan Liu pada citra digital. Pada tahap pertama, citra digital terlebih dahulu akan dibagi menjadi blok citra sesuai ukuran blok yang diinginkan. Setiap blok dilakukan perhitungan similaritas dengan menggunakan *euclidean distance* dan menghasilkan *ranking* blok citra dengan nilai similaritasnya. Pengguna kemudian menyeleksi blok yang tidak relevan salah satunya dengan melihat posisi *ranking* dari blok citra. *Feedback* dari pengguna kemudian menjadi masukan blok citra relevan yang menjadi input pada tahap kedua. Pada tahap kedua, blok citra yang relevan menurut pengguna kemudian dibagi lagi menjadi blok yang lebih kecil kemudian dilakukan proses yang sama seperti tahap pertama.

2. Research Method

2.1. Evaluasi Semi Otomatis berbasis Similaritas

Hankins dan Liu (2014) memperkenalkan sebuah metode semi otomatis untuk melakukan evaluasi berbasis similaritas. Dalam metode tersebut terdapat 2 (dua) tahapan pemrosesan yang dapat dilihat pada Gambar 1.



Gambar 1 Flowchart proses evaluasi semi otomatis berbasis similaritas[4]

Proses yang terdapat pada tahapan pertama adalah melakukan evaluasi *substring* dan melakukan perangkingan berdasarkan tingkat kemiripan dengan *substring* yang dibandingkan ($D(x,y)$). Pengukuran tingkat kemiripan antara *substring* x dan y dapat dihitung menggunakan Persamaan (1) dimana fungsi K merupakan Kolomogorov Complexity. Nilai $D(x,y)$ berada pada range 0-1 dimana semakin besar nilai $D(x,y)$ berarti *substring* x semakin mirip atau relevan terhadap *substring* y .

$$D(x, y) = 1 - \frac{K(x)+K(x|y)}{K(xy)} \tag{1}$$

Setelah diperoleh list *substring* dan nilai kemiripannya, kemudian dilakukan pertanyaan kepada pengguna untuk melakukan penilaian apakah *substring* tersebut relevan atau tidak. Jika *substring* tersebut relevan maka *substring* tersebut akan menjadi output dari tahap pertama atau input untuk tahap kedua. Algoritma untuk tahap pertama dapat dilihat pada Gambar 2.

Algorithm 1 Calculate an ordered list of values and substrings sorted by value

```

1: for all substrings  $s_x$  of length  $l/n$  in a string  $s$ , where  $l$ 
   is the length of  $s$  and  $n \leq l$  where  $n \leq l$  and  $x = 1, \dots, n$ ;
   do
2:   append  $D(s_x, v)$  to array value
3:   append  $x$  to array string
4: end for
5: sort value and string simultaneously by each element in
   value, in a descending order.

```

Gambar 2 Algoritma pada tahap pertama proses evaluasi berbasis similaritas[4]

Pada tahap kedua, setiap *string* yang merupakan *output* dari tahap pertama kemudian dibagi menjadi 4 (empat) *substring*. Setelah dilakukan pembagian menjadi 4 (empat) *substring*, kemudian dilanjutkan kembali dengan melakukan evaluasi dengan menggunakan algoritma pada tahap pertama. Proses pada tahap kedua berhenti apabila mencapai kondisi dimana *substring* yang telah dibagi tidak mengalami peningkatan nilai kesamaan yang signifikan atau ketika *substring* yang telah dibagi 4 (empat) bernilai 0 (nol). Algoritma pada tahap kedua dapat dilihat pada Gambar 3.

Algorithm 2 Subdivide string s into four substrings, and calculate and sort the substrings by their relevance.

```

1: for  $i = 0$  to  $max\_substrings$  do
2:   calculate newvalue and newstring using  $string[x]$ 
   and length  $l/4$  (using algorithm 1)
3:   append contents of newvalue to newvalues
4:   append contents of newstring to newstrings
5: end for
6: sort newvalues and newstrings simultaneously by each
   element in newvalues, in a descending order.

```

Gambar 3 Algoritma pada tahap kedua proses evaluasi berbasis similaritas[4]

2.2. Pixel-based Image Forensic

Menurut Farid (2009), terdapat beberapa 5 (lima) cara untuk melakukan *image forensic*, salah satunya adalah dengan teknik *pixel-based* yang mendeteksi anomali statistika pada tingkat *pixel*. Proses pencarian citra yang memiliki kemiripan dapat dilakukan dengan membandingkan ciri yang dimiliki oleh citra [5]. Ciri yang dapat digunakan adalah dengan membandingkan nilai warna pada setiap pixel. Salah satu pengukuran similaritas pada level *pixel* dapat menggunakan *Euclidean distance*, baik pada citra *grayscale* [6] maupun pada citra berwarna [7]. *Euclidean distance* merupakan pengukuran *image metric* yang paling banyak digunakan karena kemudahannya implementasinya [6]. *Euclidean distance* dari citra *grayscale* x dan y ($D(x,y)$) dengan ukuran citra $M \times N$ dapat dilihat pada Persamaan (2).

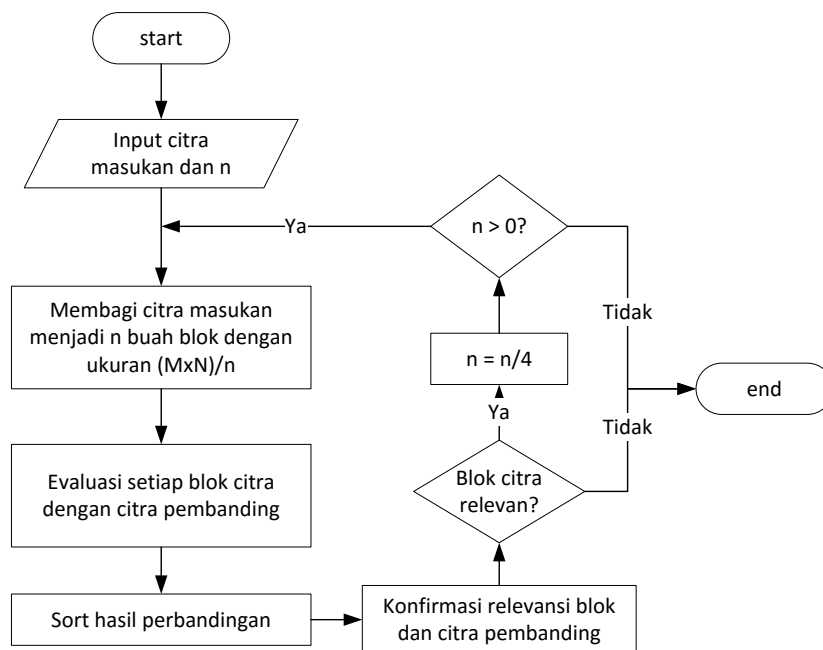
$$D(x, y) = \sqrt{\sum_{k=1}^{MN} (x_k - y_k)^2} \quad (2)$$

Pada Persamaan (2), x_k merupakan nilai keabuan pada citra x di pixel ke- k dan y_k merupakan nilai keabuan pada citra y di pixel ke- k . Pada citra yang berwarna setiap komponen pixel memiliki nilai Red, Green, dan Blue (RGB). *Euclidean distance* untuk mengukur kemiripan citra berwarna RGB x dan y dapat dilihat pada Persamaan (3) [7]. Pada Persamaan (3), x_k^R, x_k^G, x_k^B merupakan nilai warna Red, Green, dan Blue pada citra x di pixel ke- k dan y_k^R, y_k^G, y_k^B merupakan nilai warna Red, Green, dan Blue pada citra y di pixel ke- k .

$$D(x, y) = \sqrt{\sum_{k=1}^{MN} [(x_k^R - y_k^R)^2 + (x_k^G - y_k^G)^2 + (x_k^B - y_k^B)^2]} \quad (3)$$

3. Result and Analysis

Deshpande dan Kanikar (2012) memperkenalkan suatu teknik deteksi pemalsuan citra yang diberi nama *Fast Copy-Move Forgery Detection*. Pada teknik berbasis pada pixel tersebut dikhususkan untuk mendeteksi bagian dari citra yang di-copy dan disisipkan (*move*) pada citra itu sendiri [1]. Teknik tersebut dikhususkan untuk mendeteksi bagian dari citra yang sama sehingga tidak dapat digunakan untuk mendeteksi bagian pada citra yang berbeda. Keuntungan dari teknik tersebut adalah proses investigasi berlangsung secara otomatis sehingga seluruh bagian dari citra akan diproses. Kelemahan yang dapat ditemui pada metode tersebut apabila hanya bagian kecil dari gambar yang mengalami proses *copy-move*. Hal tersebut berdampak pada beberapa bagian yang tidak perlu diinvestigasi menjadi diproses sehingga berdampak pada efisiensi waktu pemrosesan. Dalam hal ini maka tetap dibutuhkan peran dari pengguna untuk melakukan seleksi terhadap bagian yang relevan dengan citra pembandingan. Kelemahan tersebut akan dicoba diatasi dengan mengkaji penerapan dari metode evaluasi semi otomatis berbasis similaritas pada *pixel-based image forensic*.



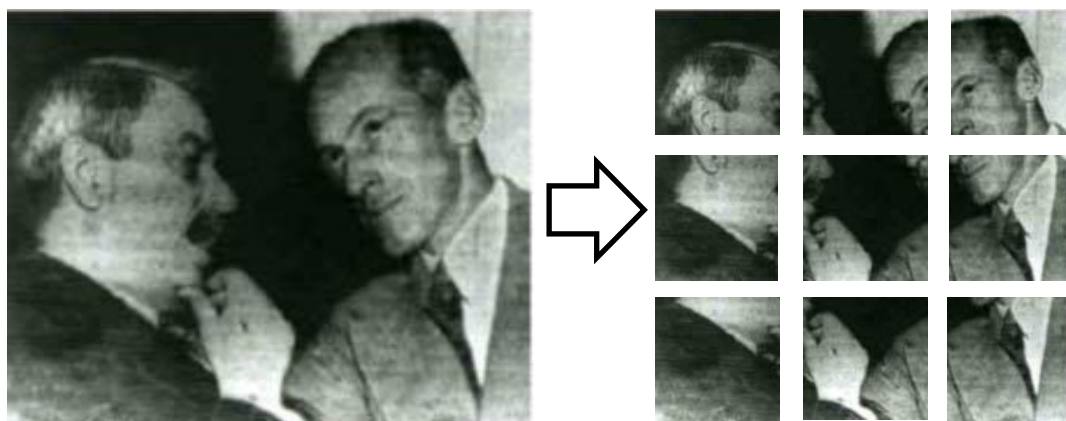
Gambar 4 Flowchart proses evaluasi semi otomatis berbasis similaritas pada data citra

Pada metode berbasis similaritas yang diperkenalkan oleh Hankins dan Liu (2014) dimana yang menjadi objek penelitian adalah *string* dan perhitungan nilai similaritas menggunakan Kolomogorov Complexity. Pada tulisan ini akan dikaji penggunaan metode yang sama tetapi dengan objek penelitian berupa data citra (Gambar 4). Representasi dari citra tersebut nantinya berupa blok citra yang terdiri dari nilai *pixel* sesuai posisi dari blok tersebut. Pengukuran tingkat kemiripan antara blok dengan setiap bagian pada citra pembandingan akan menggunakan *Euclidean distance* dimana semakin kecil nilai *Euclidean distance* maka semakin mirip blok citra yang diperbandingkan.



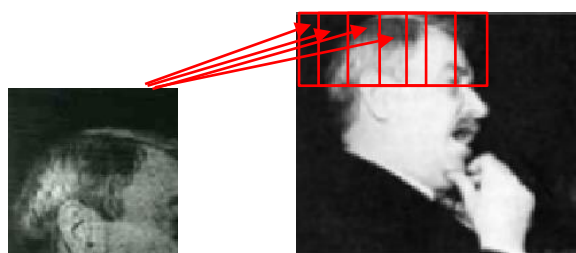
Gambar 4 Citra yang merupakan penggabungan dari dua citra asli[9]

Pada tahap pertama, data citra akan dibagi menjadi n bagian atau blok, dimana setiap blok memiliki ukuran yang sama yaitu $(M \times N)/n$ (Gambar 6). Setiap blok kemudian akan dihitung dengan similaritas menggunakan *Euclidean distance* dengan setiap bagian pada citra pembandingan sesuai ukuran blok citra input (Gambar 7). Hasil pengukuran setiap blok dengan setiap bagian pada citra pembandingan kemudian dibentuk *ranking* secara *ascending*. Blok yang memiliki nilai *Euclidean distance* paling kecil berarti blok tersebut memiliki kemiripan yang tinggi dengan citra pembandingan. Setelah dibentuk *ranking* terhadap setiap blok, kemudian dilakukan konfirmasi kepada pengguna apakah blok tersebut relevan dengan citra pembandingan atau tidak. Apabila blok tersebut dinilai relevan maka blok tersebut akan menjadi input pada tahap kedua. Sebaliknya jika tidak, maka blok tersebut tidak akan digunakan pada pemrosesan tahap berikutnya.



Gambar 6 Pembagian blok citra menjadi 9 blok dengan ukuran blok 150x150 pixel

Pada tahap kedua kemudian dilakukan pembagian setiap blok yang dianggap sebagai bagian citra paling relevan menjadi 4 (empat) bagian. Blok yang terbagi tersebut kemudian kembali dilakukan perhitungan ulang nilai *Euclidean distance* terhadap citra pembandingan. Setelah diperoleh nilai *Euclidean distance* yang baru pada blok kecil, kemudian kembali dibutuhkan konfirmasi kepada pengguna apakah blok tersebut relevan atau tidak. Proses pada tahapan kedua ini berlangsung secara iteratif hingga dipenuhi kondisi berhenti. Kondisi berhenti adalah ketika nilai *Euclidean distance* dari setiap blok yang dibagi mejadi 4 (empat) bagian tidak mengalami peningkatan yang signifikan atau ketika blok yang dibagi telah menghasilkan blok dengan ukuran 0 (nol) pixel.



Gambar 7 Ilustrasi pengecekan similaritas blok citra (kiri) pada citra pembandingan (kanan)

4. Conclusion

Kesimpulan yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Metode evaluasi semi otomatis berbasis similaritas tidak hanya dapat diterapkan pada data *string* tetapi juga dapat diterapkan pada data citra yang berbasis pada nilai warna pada setiap *pixel*.
2. Metode evaluasi semi otomatis berbasis similaritas pada dasarnya melakukan pembagian data ke dalam bagian besar terlebih dahulu kemudian dilanjutkan dengan pembagian ke dalam bagian yang lebih detail atau pecahan dari bagian yang besar.
3. Peranan pengguna pada metode (semi otomatis) tetap memiliki peranan penting dalam investigasi forensik terutama dalam melakukan penilaian apakah suatu bagian data relevan atau tidak. Hal tersebut dapat berdampak pada waktu pemrosesan yang dapat lebih cepat dimana bagian data yang tidak relevan tidak perlu diproses lebih lanjut.

References

- [1] Deshpande, P.; Kanikar, P.;, 2012, "Pixel Based Digital Image Forgery Detection Techniques," *International Journal of Engineering Research and Applications*, vol.2 no.3, pp. 539-543, Mei 2012
- [2] Farid, H., 2009, "Image Forgery Detection: A Survey,". *IEEE Signal Processing Magazine*, vol.no., pp. 16-25.
- [3] Charpe, J.; Bhattacharya, A.;, 2014, "Literature Survey on Forensic Techniques for Image Forgery Detection," *International Journal of Advance Research in Computer Science and Management Studies*, vol.2 no.11, pp. 510-513, November 2014
- [4] Hankis, R. Q.; Liu, J.;, 2014, "A Novel Approach to Evaluating Similarity in Computer Forensic Investigations," *Electro/Information Technology (EIT) 2014 IEEE International Conference on*, vol.no., pp. 567-572, 5-7 Juni 2014
- [5] Lisangan, E. A., 2015, "Scene Completion Menggunakan Template Matching Dan Poisson Blending," *TEMATIKA, Journal of Informatics and Information Systems*, vol.3 no.1, pp. 9-14, Maret 2015
- [6] Wang, L.; Zhang, Y.; Feng, J.;, 2005, "On the Euclidean Distance of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27 no.8, pp. 1334-1339, Agustus 2015
- [7] Sheela R. B.; Nandhitha N.M.; Manoharan, N.; Venkataraman, B.; Vasudevan, M.; Chandrasekar, Kalyana S. P.; Raj, B.; 2008, "Euclidean Distance Based Color Image Segmentation Algorithm for Dimensional Characterization of Lack of Penetration from Weld Thermographs for On-Line Weld Monitoring in GTAW," *17th World Conference on Nondestructive Testing*, vol.no., pp. -, 25-28 Oktober 2008.
- [8] Farid, H., 2008, *Digital Image Forensics*, Diunduh dari <http://www.cs.dartmouth.edu/farid/downloads/tutorials/digitalimageforensics.pdf>.