

Perbaikan Citra Dokumen Hasil Pindai Menggunakan Metode *Simple*, *Adaptive-Gaussian*, dan *Otsu Binarization Thresholding*

Dessy Tri Anggraeni

Fakultas Ilmu Komputer, Universitas Gunadarma
INDONESIA

dessytri@staff.gunadarma.ac.id

Abstract – The use of digital images from scanned documents is commonly used both for data backup and for further processing. However, often the digital image obtained is not optimum due to various factors like noise. The method to improve the quality of digital images is to filter images using the Thresholding method. This study compares three Thresholding methods, which are Simple Thresholding, Adaptive-Gaussian Thresholding, and Otsu Binarization. All three methods have advantages and disadvantages. However, using the MSE and PSNR assessment parameters, the Simple Thresholding method shows better quality with an MSE value of 5,196.76, followed by Otsu Binarization with a value of 5,934.10, and Adaptive-Gaussian Thresholding with a value of 9,025.29. Meanwhile, by using PSNR, the value for Simple Thresholding is 13.37, followed by Otsu Binarization with a value of 12.47, and Adaptive-Gaussian Thresholding with a value of 10.31.

Keyword: Adaptive-Gaussian Thresholding; Digital Image; Simple Thresholding; Thresholding; Otsu Binarization.

1. Pendahuluan

Citra adalah gambar pada bidang dwimatra (dua dimensi). Menurut Munir, citra digital adalah citra yang dihasilkan dari proses digitalisasi. Sedangkan digitalisasi adalah representasi citra dari citra malar (kontinu) menjadi nilai-nilai diskrit agar dapat diolah menggunakan komputer [1]. Menurut Sutoyo, citra digital merupakan larik (*array*) yang berisi nilai-nilai real maupun kompleks yang direpresentasikan dengan deretan bit tertentu [2]. Citra digital banyak dihasilkan dari peralatan digital seperti kamera dan scanner. Secara khusus scanner banyak digunakan untuk membuat salinan dokumen fisik menjadi dokumen digital dalam bentuk citra digital. Citra digital ini bisa digunakan untuk *backup* data maupun diolah lebih lanjut seperti pada kasus konversi citra ke dokumen digital, teknologi OCR (*Optical Character Recognition*), koreksi lembar jawab otomatis, dan lain-lain.

Akan tetapi, seringkali citra digital yang dihasilkan dari hasil *scanning* memiliki banyak *noise* (gangguan) maupun memberikan hasil yang kurang optimal dikarenakan oleh berbagai faktor, seperti kurangnya pencahayaan, kondisi berkas, kualitas alat, dan lain-lain. Hal ini akan mempengaruhi output pada proses-proses pengolahan citra selanjutnya. Citra digital yang demikian perlu diperbaiki agar dapat memberikan hasil yang optimal.

Salah satu metode perbaikan citra yang dapat digunakan adalah Metode *Thresholding* (pengambangan). Secara sederhana metode ini akan memisahkan tingkat keabuan citra berdasarkan batas ambang tertentu [3].

Metode ini cocok digunakan untuk perbaikan citra dokumen dikarenakan biasanya hanya dibutuhkan dua warna yaitu warna hitam untuk isi dokumen dan warna putih untuk latar belakang. Variasi Metode *Thresholding* ini pun bermacam-macam. Pada penelitian ini akan digunakan Metode *Simple Thresholding*, *Adaptive-Gaussian Thresholding*, dan *Otsu Binarization*. Dengan digunakannya beberapa variasi metode ini diharapkan dapat memberikan pilihan metode terbaik yang bisa digunakan untuk perbaikan kualitas citra dokumen hasil pindai.

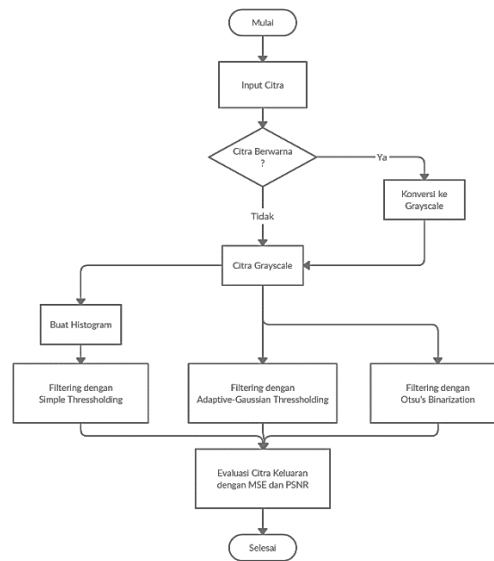
2. Metodologi

Metode penelitian merupakan gambaran tahapan-tahapan yang dilakukan dalam penelitian sehingga pelaksanaan penelitian dapat berjalan dengan lebih terstruktur. Gambar 1 berikut merupakan metodologi dalam penelitian ini.

Penelitian diawali dengan memasukkan input citra digital. Citra digital yang dimasukkan bisa berupa citra berwarna maupun *grayscale* dengan format JPEG atau JPG.

Langkah selanjutnya adalah mengubah citra inputan menjadi citra *grayscale*. Jika citra inputan adalah citra berwarna, maka harus dikonversi dahulu menjadi citra *grayscale*. Sedangkan jika citra inputan sudah dalam bentuk citra *grayscale*, maka tidak dilakukan proses konversi.

Setelah itu dilakukan proses *filtering* citra menggunakan metode-metode *Thresholding* yang telah ditentukan yaitu *Simple Thresholding*, *Adaptive-Gaussian Thresholding*, dan *Otsu Binarization*.



Gambar 1. Metodologi Penelitian

Output citra yang dihasilkan bisa dinilai secara subyektif dan obyektif. Secara subyektif penilaian dilakukan dengan melihat secara langsung output citra. Sedangkan secara obyektif dilakukan dengan menggunakan beberapa parameter seperti *Mean Square Error (MSE)* dan *Peak Signal to Noise Ratio (MSE)*.

A. Metode Thresholding

Thresholding adalah salah satu metode segmentasi citra yang memisahkan antara obyek dengan *background* dalam suatu citra berdasarkan pada perbedaan kecerahannya atau gelap terangnya. *Region* citra yang cenderung gelap akan dibuat semakin gelap, sedangkan *region* citra yang cenderung terang akan dibuat semakin terang [4]. *Thresholding* digunakan untuk men-segmentasi gambar dengan pengaturan semua piksel yang nilai intensitasnya di atas ambang batas nilai menjadi latar depan dan semua piksel yang tersisa menjadi latar belakang [5].

B. Simple Thresholding

Simple Thresholding adalah metode *Thresholding* yang akan memisahkan warna pada citra menjadi dua kelas berdasarkan batas ambang tertentu (*threshold value*). Batas ambang ini akan digunakan untuk seluruh pixel di dalam citra. Jika nilai pixel lebih kecil dari batas ambang maka akan diset ke nilai bawah (nilai 0), sedangkan jika nilai pixel lebih besar dari batas ambang maka akan diset ke nilai atas (nilai 255) [5]. Nilai batas ambang bisa ditentukan dengan melihat histogram citra. Nilai batas ambang dipilih untuk membuat galat sekecil mungkin. Nilai batas ambang adalah *edge* atau siku dari garis diagram pada histogram [1].

C. Adaptive-Gaussian Thresholding

Adaptive Thresholding adalah metode *Thresholding* yang dilakukan dengan memecah citra menjadi *region-region* yang lebih kecil. Selanjutnya setiap *region* akan memiliki

nilai batas ambang masing-masing yang akan digunakan untuk melakukan proses *Thresholding* [1]. Metode ini dianggap dapat mempertahankan lebih banyak informasi dibandingkan dengan *simple Thresholding*. Nilai ambang ditentukan dengan menggunakan metode *Gaussian* [6].

Metode *Adaptive Threshold* cocok digunakan sebagai ambang batas pada citra yang memiliki kerapatan nilai yang hampir sama [7].

D. Otsu Binarization

Metode Otsu bertujuan untuk membagi histogram citra keabuan kedalam dua daerah yang berbeda secara otomatis tanpa adanya bantuan dari pengguna untuk memasukkan nilai ambang. Pendekatan yang dilakukan oleh metode otsu yaitu dengan analisis diskriminan yang menentukan suatu variabel sehingga dapat membedakan antara dua atau lebih kelompok yang muncul secara alami. Untuk memisahkan objek dengan latar belakang analisis diskriminan akan memaksimalkan variabel tersebut [8].

Menurut Debi Razabni Erika dan Sinar Sinurat (2020), metode Otsu Binarization dapat mendeteksi suatu citra digital walaupun memiliki tingkat *noise* yang tinggi [9].

E. Metode Evaluasi MSE

Mean Square Error (MSE) adalah parameter yang sering digunakan sebagai indikator untuk mengukur kemiripan dua buah citra. Parameter ini menghitung nilai error kuadrat rata-rata dari citra asli dengan citra hasil pengolahan [10]. Semakin kecil nilainya maka kedua citra dikatakan semakin mirip. Rumus (1) perhitungan MSE adalah sebagai berikut:

$$MSE = \frac{1}{m \times n} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} [f(i, j) - g(i, j)]^2 \quad (1)$$

Dimana pada rumus (1); $m \times n$ adalah ukuran panjang x lebar citra, $f(i, j)$ adalah titik koordinat pixel citra asli, dan $g(i, j)$ adalah titik koordinat pixel citra hasil pengolahan.

F. Metode Evaluasi PSNR

Peak Signal to Noise Ratio (PSNR) adalah perbandingan antara nilai maksimum dari sinyal yang diukur dengan besarnya derau yang berpengaruh pada sinyal tersebut [11]. Untuk menghitung nilai PSNR perlu dihitung terlebih dahulu nilai MSE-nya. Satuan PSNR adalah db. Kedua citra dikatakan mirip apabila nilai PSNR tinggi. Rumus (2) perhitungan PSNR adalah sebagai berikut:

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \quad (2)$$

3. Hasil dan Pembahasan

Tahapan penelitian dilakukan sesuai dengan alur yang telah dibuat pada bagian metodologi penelitian.

A. Input Citra Digital

Tahap pertama adalah menentukan citra digital yang akan diperbaiki kualitasnya. Format citra ditentukan yaitu citra berekstensi .jpg atau .jpeg. Jumlah citra yang akan diujikan pada penelitian ini berjumlah 15 citra. Gambar 2 merupakan salah satu contoh citra yang akan digunakan.

B. Konversi Citra

Jika citra inputan adalah citra berwarna, maka proses pada tahap kedua adalah mengkonversi citra iputan menjadi citra *grayscale*. Namun apabila citra inputan sudah merupakan citra *grayscale*, maka bisa langsung dilanjutkan ke proses filterisasi pada tahap selanjutnya. Contoh citra *grayscale* hasil konversi dari Gambar 2 adalah seperti pada Gambar 3.

C. Filtering Citra

Filtering citra akan dilakukan menggunakan tiga metode yaitu *Simple Thresholding*, *Adaptive-Gaussian Thresholding*, dan *Otsu Binarization*.

1) Simple Thresholding

Pada metode *Simple Thresholding* perlu ditentukan nilai ambang secara manual. Nilai ambang ditentukan dengan melihat histogram citra. Gambar 3 memiliki histogram seperti ditunjukkan pada Gambar 4.

Histogram yang dihasilkan memiliki pola dua bukit, sehingga nilai ambang bisa ditentukan dengan cara melihat lembah terdalam diantara dua bukit. Dengan demikian, bisa ditentukan nilai ambang yang tepat adalah kurang lebih 170.

Nilai ambang kemudian diterapkan pada filterisasi citra hingga menghasilkan citra seperti pada Gambar 5. *Output* citra terlihat cukup bagus dimana teks/tulisan sebagian besar bisa terbaca dengan jelas. Beberapa bagian teks terlihat terlalu tebal, seperti pada tulisan "MEMBANGUN BANGSA" sehingga agak sulit teridentifikasi. *Noise* yang dihasilkan juga relatif sedikit.

2) Adaptive-Gaussian Thresholding

Pada metode *Adaptive-Gaussian Thresholding*, citra dipecah menjadi *region-region*. Nilai batas ambang dan filterisasi dilakukan untuk setiap *region*. Penentuan batas ambang ditentukan menggunakan metode Gaussian. Hasil output citra bisa dilihat pada Gambar 6.

Pada hasil terlihat bahwa *noise* yang dihasilkan cukup banyak sehingga mengganggu identifikasi teks/tulisan. Akan tetapi secara kasat mata, semua tulisan/teks tetap bisa dibaca dengan jelas. Tidak ada teks/tulisan yang terlalu tebal seperti pada metode *Simple Thresholding*.

3) Otsu Binarization Thresholding

Metode *Otsu Binarization* hampir sama dengan *Simple Thresholding*, dimana filter dilakukan untuk keseluruhan citra dengan mengacu pada nilai ambang tertentu. Akan tetapi pada *Otsu Binarization*, nilai ambang ini akan ditentukan secara otomatis oleh program menggunakan perhitungan formulasi Otsu. Output citra yang dihasilkan bisa dilihat pada Gambar 7.

Citra output hampir sama dengan output menggunakan *Simple Thresholding*. Keseluruhan teks/tulisan dapat terbaca dengan jelas. Pada bagian tulisan "MEMBANGUN BANGSA" pun, meskipun lebih tebal, akan tetapi lebih baik dibandingkan metode *Simple Thresholding*. *Noise* yang dihasilkan juga relatif sedikit.

Hasil output hasil filtering lainnya bisa dilihat pada Gambar 8 dan Gambar 9.

Pada Gambar 8 secara kasat mata terlihat bahwa output dari metode *Simple Thresholding* tidak bisa terbaca karena banyak *noise*, output dari metode *Adaptive-Gaussian* juga tidak bisa terbaca dikarenakan teks menjadi *blur*, dan *output* dari metode *Otsu Binarization* lebih bisa terbaca.

Pada Gambar 9 hasilnya sedikit berbeda dimana output dengan metode *Simple Thresholding* dan *Otsu Binarization* hampir sama dan sebagian besar bisa terbaca meskipun ada beberapa bagian yang tidak bisa dibaca. Sedangkan output metode *Adaptive-Gaussian* bisa terbaca di semua bagian.

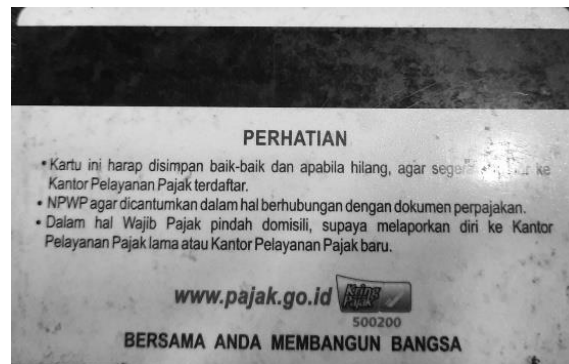
D. Evaluasi Citra Input

Evaluasi citra dihitung dengan menggunakan parameter MSE dan PSNR. Tabel 1 menunjukkan nilai MSE dan PSNR citra menggunakan masing-masing metode *Thresholding*.

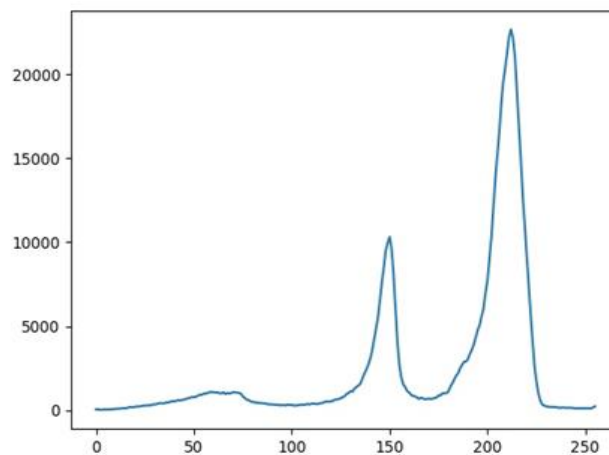
Rata-rata nilai MSE untuk *Simple Thresholding* adalah 5.196,76, *Adaptive-Gaussian Thresholding* adalah 9.025,29, dan *Otsu Binarization* adalah 5.934,10. Sedangkan rata-rata nilai PSNR untuk metode *Simple Thresholding* adalah 13.37, *Adaptive-Gaussian Thresholding* adalah 10.31, dan *Otsu Binarization* adalah 12,47.



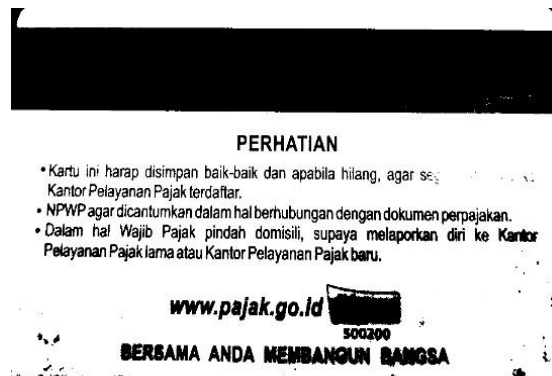
Gambar 2. Contoh Dokumen Hasil Pindai



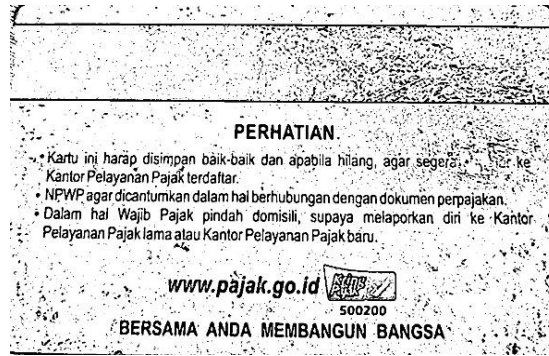
Gambar 3. Hasil Konversi ke Citra *Grayscale*



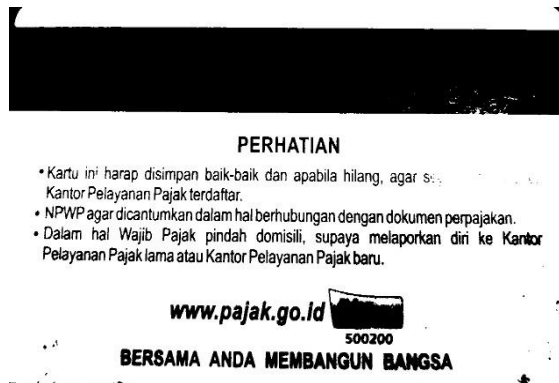
Gambar 4. Histogram Citra



Gambar 5. Citra Hasil Filterisasi dengan Metode *Simple Thresholding*



Gambar 6. Citra Hasil Filterisasi dengan Metode *Adaptive-Gaussian Thresholding*



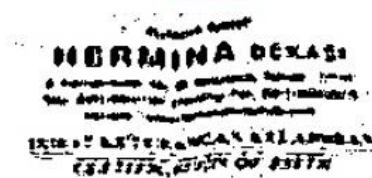
Gambar 7. Citra Hasil Filterisasi dengan Metode *Otsu Binarization Thresholding*



Gambar Asli



Simple Thresholding



Adaptive-Gaussian



OTSU Binarization

Gambar 8. Perbandingan Citra Asli dan Citra Hasil Filterisasi

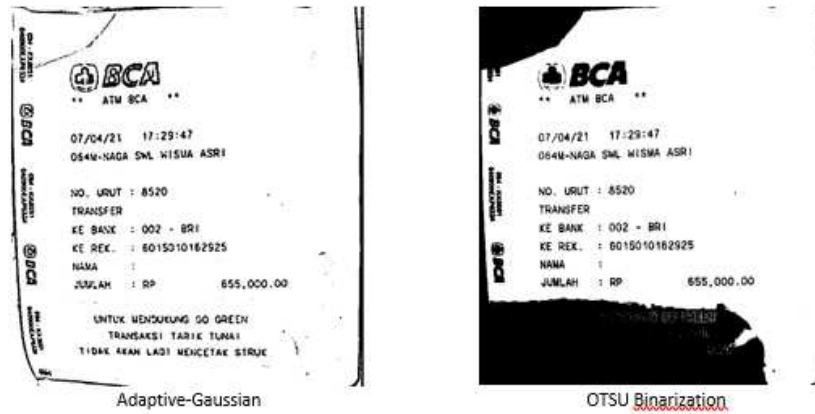


Gambar Asli



Simple Thresholding

Gambar 9. Perbandingan Citra Asli dan Citra Hasil Filterisasi (a)



Gambar 9. Perbandingan Citra Asli dan Citra Hasil Filterisasi (b)

Tabel 1. Nilai MSE dan PSNR Citra Hasil Filterisasi

Citra	MSE			PSNR		
	Simple	AG	OTSU	Simple	AG	OTSU
Citra 1	4.564,66	7.136,35	6.741,74	11,45	9,58	9,84
Citra 2	893,03	1.573,66	972,10	18,62	16,16	18,25
Citra 3	2.290,21	8.448,42	2.514,48	14,53	8,86	14,12
Citra 4	10.549,97	16.141,41	11.862,99	7,90	6,05	7,38
Citra 5	12.805,43	14.430,75	13.634,47	7,05	6,53	6,78
Citra 6	415,31	1.437,28	594,66	21,94	16,55	20,38
Citra 7	12.247,78	13.618,66	12.250,85	7,25	6,78	7,24
Citra 8	800,68	1.466,03	1.015,06	19,09	16,46	18,06
Citra 9	3.840,24	13.824,31	3.957,11	12,28	6,72	12,15
Citra 10	1.211,52	4.836,09	1.487,40	17,29	11,28	16,40
Citra 11	1.095,41	3274,53	3.941,69	17,70	12,97	12,17
Citra 12	12.702,64	15.898,42	15.381,85	7,09	6,11	6,26
Citra 13	906,03	1.429,73	983,92	18,56	16,58	18,20
Citra 14	5.174,47	14.669,32	5.219,20	10,90	6,45	10,95
Citra 15	8.453,96	11.443,60	8.453,92	8,86	7,54	8,86
Rata-rata	5.196,76	9.025,29	5.934,10	13,37	10,31	12,47

4. Kesimpulan

Penelitian ini membandingkan kinerja dari tiga metode *Thresholding* yaitu *Simple Thresholding*, *Adaptive-Gaussian Thresholding*, dan *Otsu Binarization* untuk filterisasi citra hasil pindai dokumen. Hasil evaluasi menggunakan parameter penilaian MSE dan PSNR menunjukkan bahwa metode *Simple Thresholding* adalah yang terbaik. Dari 15 uji coba yang dilakukan, rata-rata nilai MSE untuk *Simple Thresholding* adalah 5.196,76, diikuti *Otsu Binarization* dengan nilai 5.934,10, dan *Adaptive-Gaussian Thresholding* dengan nilai 9.025,29. Begitu pula dengan menggunakan parameter PSNR, metode *Simple Thresholding* juga menunjukkan hasil yang terbaik dengan rata-rata nilai PSNR 13.37, diikuti *Otsu Binarization* dengan nilai 12,47, dan *Adaptive-Gaussian Thresholding* dengan nilai 10,31.

5. Daftar Pustaka

- [1] R. Munir, Pengolahan Citra Digital dengan Pendekatan Algoritmik, Bandung: Informatika, 2004.
- [2] D. Putra, Pengolahan Citra Digital, Yogyakarta: Penerbit Andi, 2010.
- [3] D.-H. P. a. S. Meignen, "A Novel Thresholding Technique for the Denoising of Multicomponent Signals," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 4004–4008, 2008.
- [4] W. D. H. A. S. Setiawan I, "Pengolah Citra dengan Metode Thresholding dengan Matlab R2014A," *Jurnal Media Infotama*, vol. 15, no. 2, 2019.
- [5] V. S. N. R. Dnyandeo, "Review of Adaptive Thresholding Techniques for Vehicle Number Plate Recognition," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 4, 2016.
- [6] OpenCV, "Transformasi Citra Biner Menggunakan Metode Thresholding Dan Otsu

- Thresholding," [Online]. Available: https://docs.opencv.org/master/d7/d4d/tutorial_py_thresholding.html. [Accessed 2021].
- [7] I. G. S. Moch Hatta, "Counting Sperma Aktif Menggunakan Metode Otsu Threshhold dan Local Adaptive Threshhold," *Teknika : Engineering and Sains Journal*, vol. 1, no. 1, pp. 47-54, 2017.
- [8] P. D, "Binerisasi Citra Tangan Dengan Metode Otsu," *Majalah Ilmiah Teknologi Elektro Universitas Udayana*, vol. 3, no. 2, 2004.
- [9] S. S. Debi Razabni Erika, "Analisa dan Perbandingan Algoritma Otsu Thresholding dengan Algoritma Region Growing Pada Segmentasi Citra Digital," *Journal of Computer System and Informatics (JoSYC)*, vol. 2, no. 1, pp. 9-16, 2020.
- [10] W. O. H. R. Suia D, "Analisis Kualitas Citra Medis Terkompreso JPEG," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 2, 2019.
- [11] Y. W. d. Syefrida, "Algoritma Penyisipan Frame Untuk Peningkatan Akurasi Metode Aligned Peak Signal-to-Noise Ratio Dalam Pengukuran Kualitas Video," *Jurnal Komputer Terapan*, vol. 1, no. 2, pp. 45-56, 2015.