Jurnal Teknologi Informasi dan Multimedia

e-ISSN: 2684-9151
p-ISSN: 2715-2529

https://journal.sekawan-org.id/index.php/jtim

Pelokalan Manipulasi Citra Digital Menggunakan *Block-Based Fragile Watermarking* LSB dan Fungsi Hash Kriptografi

Raihan Budi Waluyo[1], Ermadi Satria Wijaya[1*], Sigit Sugiyanto[1], Harjono[1]

[1] Program Studi Teknik Informatika, Universitas Muhammadiyah Purwokerto, Indonesia
[*] Corresponding author: juniorrehan27@gmail.com

Sitasi: R. B. Waluyo, E. S. Wijaya, S Sugiyanto, and H. Harjono, "Pelokalan Manipulasi Citra Digital Menggunakan *Block-Based Fragile Watermarking* LSB dan Fungsi Hash Kriptografi," *Jurnal Teknologi Informasi dan Multimedia*, vol. 8, no. 4, pp. 817–829, 2026.

Diterima : 29-07-2026
Direvisi : 05-08-2026
Disetujui : 08-08-2026
Diterbitkan : 02-09-2026

Abstrak. Meningkatnya kemudahan akses terhadap perangkat lunak penyuntingan citra telah menjadikan pemalsuan citra digital sebagai persoalan yang semakin mengkhawatirkan, khususnya pada bidang forensik, dokumen legal, dan pencitraan medis yang menuntut keaslian konten. Penelitian ini mengusulkan skema *fragile watermarking* berbasis matriks blok 2×2 piksel yang tidak saling tumpang tindih (*non-overlapping*), diintegrasikan dengan fungsi hash kriptografi MD5 untuk mencapai pelokalan manipulasi citra digital secara presisi. Penyisipan *watermark* dibatasi secara eksklusif pada ruang 1-bit *Least Significant Bit* (LSB), sementara 7-bit *Most Significant Bit* (MSB) dari setiap blok diproses menggunakan MD5, dipotong menjadi 4-bit, dan disisipkan melalui operasi logika OR. Skema ini diuji pada 30 sampel dari dataset pemalsuan CoMoFoD melalui tahap penyisipan *watermark*, simulasi serangan manipulasi spasial, ekstraksi buta, dan pelokalan, dengan kinerja diukur menggunakan PSNR, SSIM, dan akurasi deteksi. Hasil menunjukkan rata-rata PSNR sebesar 51,09 dB dan SSIM sebesar 0,9977, yang mengindikasikan tidak adanya degradasi visual yang dapat dideteksi, serta akurasi deteksi rata-rata sebesar 93,51%, yang mengonfirmasi sensitivitas tinggi terhadap manipulasi spasial dengan tingkat deteksi palsu yang minimal. Temuan ini menunjukkan bahwa pendekatan hashing berbasis blok mikro secara efektif menyeimbangkan ketidaktampakan visual dan kerapuhan, sehingga menjadi instrumen forensik citra digital yang ringan secara komputasi namun tetap presisi.

Kata kunci: fragile watermarking; hash MD5; pelokalan manipulasi; forensik citra; autentikasi citra digital

Abstract. The increasing accessibility of digital image editing tools has made image forgery a growing concern in forensic, legal, and medical imaging contexts, where content authenticity is critical. This study proposes a fragile watermarking scheme based on non-overlapping 2×2 pixel block matrices integrated with the MD5 cryptographic hash function to enable precise tamper localization in digital images. Watermark embedding is restricted exclusively to the 1-bit Least Significant Bit (LSB) space, while the remaining 7-bit Most Significant Bit (MSB) of each block is processed using MD5, truncated to 4 bits, and embedded through a bitwise OR operation. The scheme was evaluated on 30 samples from the CoMoFoD forgery dataset through watermark embedding, spatial manipulation attack simulation, blind extraction, and localization, with performance measured using PSNR, SSIM, and detection accuracy. The results show an average PSNR of 51.09 dB and SSIM of 0.9977, indicating no perceptible visual degradation, alongside an average detection accuracy of 93.51%, confirming high sensitivity to spatial tampering with minimal false positives. These findings demonstrate that micro block-level hashing effectively balances imperceptibility and fragility, offering a computationally lightweight yet precise instrument for digital image forensics.

Keywords: fragile watermarking; MD5 hash; tamper localization; image forensics; digital image authentication

1. Pendahuluan

Citra digital telah menjadi salah satu bentuk informasi visual yang paling banyak diproduksi, didistribusikan, dan dikonsumsi pada era masyarakat digital saat ini, mulai dari dokumentasi jurnalistik, bukti forensik hukum, hingga rekam medis digital. Kemudahan akses terhadap perangkat lunak penyuntingan citra telah menurunkan secara drastis hambatan teknis untuk melakukan manipulasi konten, sehingga keaslian sebuah citra tidak lagi dapat diasumsikan hanya berdasarkan tampilan visualnya semata. Kondisi ini menempatkan isu integritas dan autentikasi citra digital sebagai persoalan yang mendesak untuk ditangani, terutama pada domain yang menuntut akurasi tinggi seperti forensik digital, verifikasi dokumen legal, dan pencitraan medis, di mana satu piksel yang termanipulasi dapat berdampak pada keputusan hukum maupun klinis.

Urgensi persoalan ini tercermin dari maraknya kasus manipulasi foto digital yang menjadi objek kajian forensik di Indonesia dalam beberapa tahun terakhir, mulai dari pemalsuan foto pada media sosial [1] hingga kasus serupa pada bukti digital dalam proses peradilan[2]. Berbagai instansi penegak hukum dan peneliti forensik digital di Indonesia secara konsisten mengembangkan metode analisis keaslian citra, seperti pendekatan berbasis *Error Level Analysis* (ELA)[3]. Pendekatan tersebut juga dipadukan dengan metode NIST untuk menelusuri jejak manipulasi pada citra digital sebagai barang bukti[4]. Tren ini menegaskan bahwa ancaman terhadap keaslian citra digital bukan lagi isu marjinal, melainkan tantangan keamanan informasi berskala luas yang membutuhkan mekanisme deteksi yang andal, presisi, dan idealnya dapat diverifikasi secara mandiri tanpa bergantung pada citra asli sebagai pembanding. Signifikansi persoalan ini turut didukung oleh data kepolisian nasional, yang mencatat kasus pemalsuan dokumen dan citra digital berada pada kategori pelaporan tertinggi sepanjang tahun 2021, dengan potensi kerugian materiil maupun immateril bagi korban[5].

Sejauh ini, penelitian deteksi pemalsuan citra di Indonesia sebagian besar berkembang pada pendekatan forensik pasif (*passive forensics*), yaitu menganalisis jejak manipulasi setelah citra beredar. Salah satu pendekatan yang digunakan adalah pencocokan blok (*block matching*) untuk mendeteksi *copy-move*[6]. Pendekatan lain memanfaatkan analisis koefisien *Discrete Cosine Transform* dengan metode ordinal measure[7]. Algoritma *template matching* juga digunakan untuk memverifikasi foto di media sosial[1]. Pada sisi lain, pendekatan aktif berbasis *watermarking* dan fungsi hash kriptografis juga telah dieksplorasi sebagai instrumen autentikasi citra. Salah satu contohnya adalah penyisipan *watermark* yang memanfaatkan kombinasi *pixel value indicator* dan *most significant bit*[8]. Pendekatan lain menerapkan deteksi keaslian uang kertas berbasis *watermark* pada citra digital[9]. Terdapat pula penelitian yang membandingkan kinerja *Discrete Cosine Transform* dan *Discrete Wavelet Transform* untuk *watermarking* pada citra RGB[10]. Selain itu, fungsi hash MD5 telah dimanfaatkan untuk menguji orisinalitas citra digital[11]. Pendekatan serupa juga diterapkan menggunakan fungsi hash SHA-256[12]. Pada level internasional, perkembangan skema fragile *watermarking* berbasis *self-embedding* juga banyak dikaji. Salah satu kajiannya berupa tinjauan sistematis atas mekanisme penyisipan dan pelokalan kerusakan[13]. Kajian lain mengusulkan skema berbasis *Absolute Moment Block Truncation Coding* (AMBTC) dan *Vector Quantization* (VQ) untuk pelokalan dan pemulihan citra[14]. Terdapat pula skema *self-recovery* yang dirancang khusus untuk citra medis[15].

Meskipun demikian, tinjauan terhadap literatur di atas menunjukkan bahwa masih terdapat kesenjangan yang signifikan. Pertama, mayoritas skema forensik citra yang dikembangkan di Indonesia masih bersifat pasif dan deskriptif dengan mengandalkan alat bantu eksternal seperti ELA[3]. Pendekatan lain mengandalkan analisis metadata menggunakan metode NIST[4]. Skema lain juga menggabungkan ELA dengan *clone detection* dan *Exif* untuk mendeteksi keaslian gambar[16]. Skema-skema tersebut tidak menyediakan mekanisme autentikasi

aktif yang dapat memverifikasi integritas citra tanpa citra pembanding. Kedua, skema *watermarking* berbasis kombinasi pixel value *indicator* dan *most significant bit* masih beroperasi pada satuan blok yang relatif besar sehingga area manipulasi hanya dapat dideteksi secara kasar dan kurang mampu memetakan batas tepi (*edge*) objek yang dimanipulasi secara akurat[8]. Keterbatasan serupa juga ditemukan pada skema berbasis hash MD5, yang belum secara khusus dirancang untuk pelokalan manipulasi hingga level piksel yang presisi[11]. Keterbatasan yang sama juga berlaku pada skema berbasis hash SHA-256[12]. Ketiga, pengujian kinerja algoritma pada dataset forensik standar yang menyediakan citra asli, citra termanipulasi, dan masker biner sebagai ground truth masih jarang dikombinasikan secara konsisten dengan skema berbasis *block matching* untuk deteksi *copy-move*[6]. Kombinasi serupa juga jarang diterapkan pada skema berbasis *ordinal measure* dari koefisien Discrete Cosine Transform[7]. Skema *watermarking* berbasis perbandingan *Discrete Cosine Transform* dan *Discrete Wavelet Transform* pada citra RGB juga belum diuji secara konsisten menggunakan dataset semacam itu, sehingga sulit dilakukan pengukuran akurasi pelokalan secara kuantitatif dan objektif[10].

Berdasarkan kesenjangan yang telah diuraikan, penelitian ini merumuskan permasalahan sebagai berikut: (1) sejauh mana penurunan ukuran blok pelokalan menjadi matriks mikro 2×2 piksel dapat meningkatkan presisi deteksi manipulasi spasial dibandingkan skema *watermarking* berbasis blok besar yang telah ada; dan (2) sejauh mana penyisipan segel autentikasi hash MD5 yang dibatasi eksklusif pada ruang 1-bit LSB mampu mempertahankan kualitas visual (*imperceptibility*) citra sekaligus tetap sensitif terhadap manipulasi sekecil apa pun (*fragility*).

Berangkat dari kesenjangan tersebut, penelitian ini mengusulkan skema *fragile watermarking* dengan resolusi blok mikro berukuran 2×2 piksel yang diintegrasikan dengan fungsi hash MD5, dengan penyisipan segel autentikasi secara eksklusif pada ruang 1-bit *Least Significant Bit* (LSB) tanpa mengganggu struktur 7-bit *Most Significant Bit* (MSB). Pendekatan ini dirancang untuk mengeksplorasi sejauh mana granularitas blok yang sangat kecil dapat meningkatkan presisi pelokalan manipulasi spasial tanpa mengorbankan kualitas visual citra, dengan pengujian dilakukan secara sistematis pada dataset forgery standar melalui tahapan penyisipan *watermark*, simulasi serangan manipulasi, ekstraksi buta (*blind extraction*), dan evaluasi kinerja menggunakan metrik PSNR, SSIM, dan akurasi deteksi.

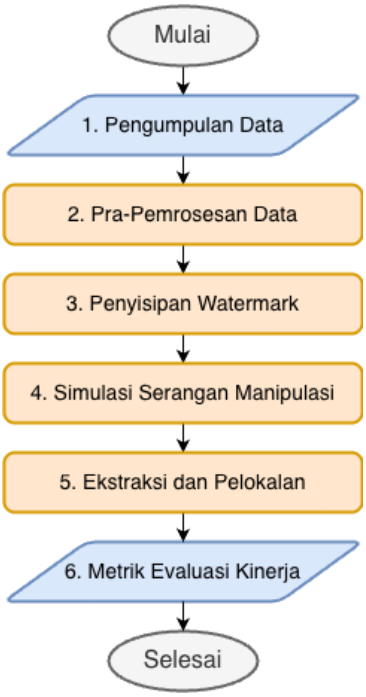
Pemilihan ukuran blok 2×2 piksel, alih-alih blok berukuran lebih besar seperti pada skema berbasis pixel value indicator dan MSB[8], didasarkan pada pertimbangan bahwa blok 2×2 merupakan satuan terkecil yang masih memiliki empat piksel MSB untuk digabungkan sebagai masukan fungsi hash, sekaligus menyediakan tepat empat ruang LSB (satu per piksel) sebagai wadah penyisipan 4-bit keluaran hash yang terpotong. Konfigurasi ini memaksimalkan granularitas pelokalan hingga ke level piksel-piksel yang berdekatan, sehingga batas tepi (*edge*) area manipulasi dapat dipetakan jauh lebih presisi dibandingkan blok berukuran lebih besar yang hanya mampu menandai manipulasi secara kasar per kelompok piksel. Solusi yang diusulkan ini secara langsung membangun di atas dua garis penelitian yang telah dipaparkan sebelumnya, yaitu efektivitas fungsi hash MD5 dan SHA-256 untuk autentikasi citra [11, 12] serta keterbatasan granularitas blok pada skema *watermarking* aktif yang telah ada[8], sehingga kontribusi utama penelitian ini adalah memadukan kedua pendekatan tersebut pada resolusi blok yang jauh lebih kecil.

Dengan demikian, penelitian ini bertujuan untuk merancang dan mengevaluasi skema *fragile watermarking* berbasis blok mikro dan hash MD5 yang mampu menyeimbangkan ketidaktampakan visual (*imperceptibility*) dengan sensitivitas dan presisi pelokalan manipulasi pada level piksel. Secara teoretis, penelitian ini diharapkan memperkaya kajian mengenai efektivitas granularitas blok mikro dalam skema hashing untuk autentikasi citra digital, sekaligus memperluas literatur skema *watermarking* aktif di Indonesia yang selama ini masih didominasi pende-

katan forensik pasif[1]. Secara praktis, skema yang diusulkan berpotensi menjadi instrumen forensik digital yang ringan secara komputasi namun presisi, sehingga dapat diadopsi untuk verifikasi keaslian citra pada domain-domain kritis seperti forensik hukum, dokumen legal, dan pencitraan medis, di mana keandalan deteksi manipulasi sekecil apa pun menjadi krusial.

2. Bahan dan Metode

Penelitian ini mengusulkan skema *fragile watermarking* berbasis blok matriks 2×2 piksel dengan integrasi fungsi hash kriptografi MD5 untuk keperluan pelokalan manipulasi citra digital. Pendekatan ini dirancang guna mencapai dua tujuan utama secara simultan, yaitu sensitivitas pelokalan modifikasi spasial dan pelestarian fidelitas visual (*invisibility*). Secara operasional, metode ini membatasi penyisipan segel autentikasi secara eksklusif hanya pada ruang 1-bit *Least Significant Bit* (LSB) untuk meminimalkan degradasi visual, serta mengeksploitasi sifat avalanche effect dari algoritma MD5 untuk mendeteksi kerusakan piksel dengan tingkat presisi yang tinggi. Untuk merealisasikan arsitektur keamanan tersebut, metode penelitian dirancang secara sistematis melalui lima fase utama yang saling berkesinambungan. Fase pertama mencakup pengumpulan dan pra-pemrosesan data citra dari dataset standar forensik. Fase kedua merupakan proses penyisipan *watermark* (*embedding*) untuk memvalidasi integritas piksel. Fase ketiga adalah simulasi serangan manipulasi spasial terukur. Fase keempat merupakan tahap ekstraksi dan pelokalan untuk memetakan area kerusakan secara presisi. Fase terakhir adalah evaluasi komprehensif terhadap kinerja algoritma menggunakan parameter metrik visual dan akurasi.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Penelitian ini menggunakan dataset forensik standar *Copy-Move Forgery Detection* (CoMoFoD) yang terdiri dari 30 sampel citra berformat warna RGB 24-bit. Untuk setiap sampel, pengujian ini memanfaatkan tiga varian data utama, yaitu citra asli (_O) sebagai media utama yang akan dilindungi, citra termanipulasi (_F) sebagai sumber referensi serangan modifikasi spasial, dan citra masker biner (_B) yang bertindak sebagai ground truth atau acuan koordinat area manipulasi,

mengingat skema algoritma *watermarking* yang diusulkan beroperasi pada tingkat bit piksel tunggal secara presisi. Dataset CoMoFoD yang digunakan mengikuti spesifikasi baku yang diperkenalkan pada[17], dengan citra dikelompokkan ke dalam lima kategori manipulasi copy-move, yaitu translasi, rotasi, penyekalaan (scaling), kombinasi berganda, dan distorsi bebas, yang masing-masing dapat disertai pascapemrosesan berupa kompresi JPEG, penambahan derau, pengaburan (blurring), serta perubahan kecerahan/kontras. Tiga puluh sampel yang digunakan pada pengujian ini diberi label 001 hingga 030 sesuai urutan berkas pada dataset; ilustrasi visual pada Gambar 2 dan Gambar 3 diambil dari sampel bernama 001, sedangkan hasil metrik lengkap untuk seluruh sampel disajikan pada Tabel 1.

2.2. Pra-Pemrosesan Data

Tahap pra-pemrosesan bertujuan menstandarkan ruang warna citra sebelum operasi penyisipan tingkat bit dilakukan. Citra asli (_O) dan citra termanipulasi (_F) awalnya dibaca dalam format RGB 24-bit, sedangkan citra masker biner (_B) dibaca secara spesifik sebagai matriks *grayscale* dua dimensi. Selanjutnya, citra RGB tersebut dikonversi secara matematis menjadi matriks *grayscale* 8-bit berkanal tunggal. Transformasi eksplisit ini sangat krusial guna mencegah terjadinya anomali pergeseran warna (*color shifting*) apabila manipulasi bit *Least Significant Bit* (LSB) dilakukan secara asimetris pada format RGB. Melalui konversi ini, stabilitas dan presisi komputasi untuk operasi *bitwise* pada tahap penyisipan *watermark* dapat terjamin.

Konversi dilakukan menggunakan fungsi pustaka OpenCV (cv2.cvtColor dengan mode COLOR_BGR2GRAY), yang menerapkan formula standar luminositas ITU-R BT.601 terhadap ketiga kanal warna piksel sebagai berikut:

$$Y = 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B$$

2.3. Penyisipan watermark

Fase penyisipan bertujuan untuk menanamkan segel autentikasi ke dalam citra tanpa menimbulkan degradasi visual yang kasat mata. Mekanisme ini diawali dengan membagi citra *grayscale* menjadi sejumlah blok matriks berukuran 2×2 piksel yang tidak saling tumpang tindih (*non-overlapping*). Pada setiap piksel di dalam blok tersebut, struktur nilai desimalnya dipisahkan menggunakan operasi logika bitwise. Ruang 1-bit *Least Significant Bit* (LSB) pada posisi paling akhir dikosongkan (diubah menjadi nol) guna mengalokasikan ruang penyisipan data, menyisakan 7-bit *Most Significant Bit* (MSB) sebagai representasi murni informasi visual citra. Selanjutnya, deretan 7-bit MSB dari keempat piksel di dalam blok digabungkan dan diproses menggunakan fungsi hash kriptografi MD5. Untuk menyesuaikan dengan kapasitas matriks 2×2 yang terbatas pada 4 bit (satu bit dari masing-masing piksel), luaran hash MD5 dipotong (*truncated*) sehingga hanya menyisakan 4-bit pertamanya. Terakhir, keempat bit hash tersebut disisipkan secara sekuensial ke dalam ruang LSB yang telah dikosongkan menggunakan operasi logika OR, menghasilkan citra ber-*watermark* yang terproteksi oleh segel autentikasi unik pada setiap bloknya.

2.4. Simulasi Serangan Manipulasi

Untuk menguji tingkat sensitivitas dan akurasi algoritma dalam mendeteksi modifikasi spasial, dilakukan simulasi serangan pemalsuan citra (*image forgery*) pada citra yang telah disisipkan *watermark*. Proses simulasi ini dikontrol secara terukur menggunakan citra masker biner (_B) sebagai acuan koordinat target. Sebuah batas threshold diterapkan pada masker biner untuk mengekstraksi area manipulasi, di mana piksel bernilai 255 (putih) direpresentasikan sebagai area yang akan diserang. Berdasarkan acuan pemetaan tersebut, piksel-piksel citra ber-*watermark* yang berada tepat di dalam area target ditimpa secara langsung menggunakan nilai piksel yang bersesuaian dari citra termanipulasi (_F). Hasil akhir dari tahapan ini adalah citra terserang (attacked image) yang memuat mo-

difikasi objek asing, yang selanjutnya akan digunakan sebagai subjek uji utama pada fase evaluasi dan ekstraksi keamanan.

2.5. Ekstraksi dan Pelokalan Manipulasi

Fase ekstraksi dan pelokalan bertujuan untuk memverifikasi integritas citra secara mandiri, sekaligus memetakan koordinat area yang telah dimodifikasi. Tahapan ini diawali dengan menyiapkan sebuah matriks kosong (*Detection Map*) berskala abu-abu yang memiliki dimensi ukuran sama dengan citra uji. Citra uji tersebut kemudian dipindai dan dipecah kembali menjadi blok matriks berukuran 2×2 piksel. Pada setiap blok, sistem mengekstraksi 1-bit LSB dari keempat piksel menggunakan operasi bitwise AND untuk merekonstruksi 4-bit data yang bertindak sebagai "Hash Tersimpan". Secara komputasi paralel, sistem juga mengekstraksi 7-bit MSB dari blok yang sama, lalu menghitung ulang nilai MD5-nya untuk menghasilkan 4-bit data sebagai "Hash Baru". Integritas blok diverifikasi melalui logika komparasi sederhana: apabila Hash Tersimpan tidak identik dengan Hash Baru, maka struktur piksel pada blok tersebut dinyatakan telah dirusak atau dimanipulasi. Sistem kemudian melokalisasi kerusakan tersebut dengan memberikan nilai 255 (piksel putih) pada koordinat matriks *Detection Map* yang bersesuaian, sedangkan blok yang masih otentik tetap dibiarkan bernilai 0 (piksel hitam).

2.6. Metrik Evaluasi Kinerja

Evaluasi kinerja algoritma yang diusulkan diukur melalui dua parameter utama, yaitu kualitas visual (*fidelitas*) citra ber-*watermark* dan tingkat akurasi pelokalan manipulasi. Untuk pengukuran kualitas visual, penelitian ini menggunakan metrik *Peak Signal-to-Noise Ratio* (PSNR) dan *Structural Similarity Index Measure* (SSIM). PSNR mengukur tingkat distorsi atau derau (*noise*) yang ditimbulkan akibat proses penyisipan segel hash dengan membandingkan citra asli dan citra ber-*watermark*. Sementara itu, SSIM mengevaluasi tingkat kemiripan struktural antara kedua citra tersebut. Untuk evaluasi akurasi pelokalan, algoritma menggunakan pendekatan kalkulasi sensitivitas deteksi (*True Positive Rate*) secara piksel per piksel. Proses komputasi ini diawali dengan menerapkan batas threshold pada citra masker biner untuk menghitung total keseluruhan piksel yang terkonfirmasi sebagai area manipulasi (*Total Manipulated*). Selanjutnya, sistem mengkalkulasi jumlah piksel yang berhasil dideteksi dengan benar (*Correctly Detected*), dengan syarat piksel tersebut harus bernilai 255 (putih) pada luaran Detection Map dan sekaligus bernilai 255 pada masker biner. Nilai akurasi akhir didapatkan dengan membagi jumlah piksel yang terdeteksi secara tepat dengan total piksel termanipulasi, lalu dikalikan dengan 100%. Rumus persentase akurasi tersebut direpresentasikan sebagai berikut

$$Akurasi = \frac{\text{Jumlah Piksel Terdeteksi Benar}}{\text{Total Piksel Termanipulasi}} \times 100\%$$

3. Hasil

3.1. Pengumpulan Data

Sistem berhasil mengekstrak dan memuat 30 sampel citra dari dataset CoMoFoD secara iteratif. Setiap sampel secara konsisten terhubung dengan tiga berkas, yaitu: citra asli (_O), referensi serangan (_F), dan masker biner (_B). Citra asli dan citra referensi serangan sukses diimpor ke dalam memori sebagai matriks berdimensi tiga dengan kedalaman warna RGB 24-bit. Sebaliknya, citra masker biner diimpor secara spesifik sebagai matriks *grayscale* dua dimensi, mengingat fungsinya sebagai pemetaan koordinat area manipulasi secara biner (hitam dan putih).

3.2. Pra-Pemrosesan Data

Setelah data berhasil dikumpulkan, tahap pra-pemrosesan dieksekusi untuk mereduksi citra RGB (asli dan serangan) menjadi matriks *grayscale* 8-bit ber-

kanal tunggal menggunakan fungsi konversi warna matematis. Transformasi ini terbukti sukses menyintesis tiga kanal warna menjadi satu nilai intensitas tunggal pada rentang 0 hingga 255. Luaran matriks berkanal tunggal ini sangat krusial. Jika operasi pemotongan bit (LSB) dipaksakan pada matriks RGB, hal tersebut akan memicu anomali pergeseran warna (*color shifting*). Dengan matriks *grayscale*, kanvas komputasi dipastikan stabil untuk eksekusi logika bitwise pada tahap selanjutnya.

Gambar 2. Contoh Hasil Pra-Pemrosesan (Sampel 001). (Kiri) Citra RGB Asli, (Kanan) Citra *Grayscale*

3.3. Penyisipan Watermark

Tahap penyisipan *watermark* merupakan inti dari proteksi integritas citra yang diusulkan. Berdasarkan eksekusi fungsi `embed_watermark`, sistem berhasil mengintegrasikan segel autentikasi MD5 ke dalam setiap blok citra 2×2 piksel secara deterministik.

3.3.1. Analisis Kualitas Visual (Invisibility)

Secara visual, citra ber-*watermark* hasil komputasi tidak menunjukkan adanya degradasi kualitas atau noise yang dapat dideteksi oleh indra penglihatan manusia. Karakteristik ini dikenal sebagai sifat *imperseptibilitas* (tidak terlihat), di mana perbandingan antara citra asli dan citra ber-*watermark* tidak memperlihatkan perbedaan tekstur, kontras, maupun iluminasi. Keberhasilan ini dicapai karena sistem hanya memodifikasi 1-bit *Least Significant Bit* (LSB) pada setiap piksel. Dalam skala intensitas 8-bit (0 hingga 255), perubahan nilai LSB hanya akan menggeser intensitas piksel sebesar maksimal ±1 satuan. Karena deviasi ini berada jauh di bawah ambang batas sensitivitas mata manusia, maka perubahan tersebut secara estetik dianggap tidak ada.

3.3.2. Analisis Implementasi Logika Bitwise dan Integritas Hash

Secara komputasional, integrasi hash MD5 dan operasi bitwise dieksekusi secara presisi untuk menanamkan *watermark*. Awalnya, sistem mengosongkan ruang LSB (bit ke-0) pada setiap piksel melalui operasi logika AND dengan nilai 0xFE. Selanjutnya, digest MD5 dipotong menjadi 4-bit representatif menggunakan operasi AND (0x0F) dan disesuaikan dimensinya menjadi matriks melalui perintah `reshape((2, 2))`. Data hash ini kemudian disisipkan ke dalam ruang LSB yang telah kosong menggunakan operasi OR (|). Melalui eksekusi fungsi `calculate_block_hash` yang stabil ini, setiap blok 2×2 piksel kini menyimpan "tanda tangan" kriptografis 4-bit yang unik, yang akan berfungsi sebagai acuan utama dalam pendeteksian manipulasi pada fase ekstraksi.

Gambar 3. Komparasi Visual. (Kiri) Citra *Grayscale* Asli, (Kanan) Citra ber-*Watermark* yang tampak identik

3.4. Simulasi Serangan Manipulasi

Fase pengujian ketahanan algoritma dieksekusi melalui simulasi serangan manipulasi spasial menggunakan fungsi `simulate_attack`. Tujuan dari tahapan ini adalah untuk mereplikasi skenario *image forgery* (seperti *copy-move* atau *splicing*) di dunia nyata, dan mengamati dampak kerusakan pada segel *watermark*.

3.4.1. Akurasi Ekstraksi Koordinat Target

Untuk menjamin presisi lokasi manipulasi, sistem memanfaatkan citra masker biner (*ground truth*) sebagai acuan spasial. Secara matematis, operasi thresholding diterapkan pada nilai ambang 127 untuk mengonversi masker tersebut menjadi matriks logika (*boolean map*). Operasi ini memaksa piksel dengan intensitas di atas ambang batas menjadi bernilai 255 (putih murni) sebagai area target serangan, sedangkan intensitas lainnya ditekan menjadi 0 (hitam murni) sebagai latar belakang.

3.4.2. Penimpaan Matriks dan Degradasi Segel LSB

Berdasarkan koordinat matriks logika tersebut, nilai piksel dari citra referensi serangan ditimpakan ke atas citra ber-*watermark*. Secara visual, objek manipulasi tampak menyatu secara alami (Gambar 3). Namun, pada struktur memori citra, masuknya piksel asing ini secara otomatis menimpa dan menghancurkan "Hash Tersimpan" (4-bit MD5) pada ruang 1-bit LSB di area target. Kerusakan bit yang terisolasi inilah yang mendemonstrasikan sifat rapuh (*fragile*) sistem saat citra dimanipulasi.

3.5. Ekstraksi dan Pelokalan Manipulasi

Fase ekstraksi berfungsi sebagai sistem verifikasi buta (*blind authentication*), di mana algoritma harus mampu mendeteksi letak piksel yang rusak tanpa memerlukan referensi citra asli. Berdasarkan eksekusi fungsi `extract_and_authenticate`, algoritma berhasil mendeteksi anomali pada citra terserang secara presisi tingkat piksel.

3.5.1. Presisi Komparasi Hash MD5

Sistem berhasil mengalokasikan kanvas kosong (matriks bernilai 0 atau hitam) dengan dimensi yang sama dengan citra uji untuk dijadikan sebagai Peta Deteksi (*Detection Map*). Selanjutnya, pada setiap iterasi blok 2×2 piksel, sistem mengekstrak 1-bit LSB menggunakan operasi logika block & 1 untuk mendapatkan "Hash Tersimpan". Secara paralel, sistem memanggil kembali algoritma `calculate_block_hash` untuk membangkitkan "Hash Baru" dari sisa 7-bit MSB pada blok tersebut. Kedua nilai hash tersebut kemudian dibandingkan meng-

Gambar 4. Citra terserang (attacked image)

gunakan fungsi `np.array_equal`. Pada blok yang tidak diserang, kedua hash tersebut identik. Namun, pada area yang telah disimulasikan sebagai serangan, susunan LSB telah rusak. Sistem secara akurat mendeteksi ketidakcocokan logika ini dan secara deterministik memberikan nilai 255 (putih murni) pada indeks koordinat blok tersebut di dalam kanvas Detection Map.

3.5.2. Visualisasi Peta Deteksi

Hasil akhir komparasi tersebut diekstrak dalam bentuk matriks hitam-putih. Saat dikomparasikan secara visual, bentuk sebaran piksel putih (255) pada *Detection Map* luaran sistem sangat kongruen (sebangun) dengan wujud objek modifikasi pada citra masker biner (*ground truth*). Hal ini menegaskan bahwa algoritma berhasil mengisolasi lokasi kerusakan secara konsisten dengan minim deteksi palsu (*false positive*) di luar zona serangan.

Gambar 5. Hasil Pelokalan Sistem. (Kiri) Citra Terserang; (Tengah) Masker Biner Ground Truth; (Kanan) Detection Map luaran sistem

3.6. Metrik Evaluasi Kinerja

Untuk memvalidasi unjuk kerja algoritma secara objektif, kalkulasi metrik kinerja dieksekusi terhadap keseluruhan 30 sampel dataset uji. Pengujian ini mengevaluasi parameter *Peak Signal-to-Noise Ratio* (PSNR), *Structural Similarity Index Measure* (SSIM), dan persentase Akurasi yang secara komprehensif.

Tabel 1. Hasil pengujian kualitas visual dan akurasi pelokalan system

No	Citra	Resolusi	Ukuran Asli (KB)	Ukuran WM (KB)	PSNR (dB)	SSIM	Akurasi (%)
1	001	512×512	430.12	151.66	51.09	0.9978	90.48
2	002	512×512	477.47	168.84	51.16	0.9967	93.19
3	003	512×512	371.10	136.80	50.78	0.9974	96.85
4	004	512×512	415.33	151.63	51.14	0.9950	92.72
5	005	512×512	417.09	151.99	51.13	0.9950	93.66
6	006	512×512	416.44	151.91	51.14	0.9950	93.57
7	007	512×512	406.35	148.39	50.86	0.9976	91.84
8	008	512×512	363.68	129.09	51.16	0.9947	91.25
9	009	512×512	363.68	129.09	51.16	0.9947	92.57
10	010	512×512	372.27	143.24	50.92	0.9970	91.32
11	011	512×512	571.60	200.33	51.14	0.9992	92.94
12	012	512×512	263.50	104.91	50.49	0.9970	95.48
13	013	512×512	563.78	190.44	51.14	0.9987	93.86
14	014	512×512	543.15	187.71	51.14	0.9982	93.02
15	015	512×512	543.66	191.01	51.13	0.9986	93.19
16	016	512×512	548.11	204.42	51.15	0.9993	92.99
17	017	512×512	598.93	209.64	51.09	0.9994	97.50
18	018	512×512	618.18	214.62	51.12	0.9995	94.74
19	019	512×512	618.18	214.62	51.12	0.9995	94.32
20	020	512×512	618.18	214.62	51.12	0.9995	95.79
21	021	512×512	510.33	171.75	51.15	0.9982	92.71
22	022	512×512	511.88	175.27	51.16	0.9986	93.90
23	023	512×512	427.60	166.97	51.12	0.9965	93.47
24	024	512×512	448.00	201.39	51.14	0.9984	93.96
25	025	512×512	464.92	167.65	51.15	0.9979	92.18
26	026	512×512	470.43	160.15	51.16	0.9969	93.47
27	027	512×512	617.63	213.33	51.13	0.9995	94.52
28	028	512×512	535.85	186.80	51.14	0.9986	93.69
29	029	512×512	678.29	233.86	51.12	0.9998	93.65
30	030	512×512	475.49	151.71	51.15	0.9964	92.33
Rata-rata					**51.09**	**0.9977**	**93.51**

Berdasarkan hasil komputasi pada Tabel 1, kualitas fidelitas visual algoritma terbukti sangat unggul dengan perolehan nilai rata-rata PSNR sebesar 51,09 dB. Nilai yang berada jauh di atas ambang batas standar (40 dB) ini mengindikasikan bahwa tingkat *Mean Square Error* (MSE) mendekati nol, sehingga penyisipan 4-bit hash MD5 tidak memicu derau visual. Keutuhan struktural citra ber-*watermark* juga tervalidasi melalui nilai rata-rata SSIM sebesar 0,9977, yang merepresentasikan bahwa struktur spasial piksel tetap hampir identik dengan citra aslinya. Sebagai parameter keamanan utama, algoritma mencatatkan akurasi pelokalan rata-rata sebesar 93,51%. Persentase ini mengonfirmasi bahwa penggunaan resolusi blok berukuran mikro (2×2 piksel) memberikan sensitivitas deteksi yang sangat tajam. Sistem terbukti mampu membedakan matriks otentik dan matriks termanipulasi secara presisi hingga ke detail tepian objek terkecil, menjadikan algoritma ini sangat andal untuk diimplementasikan sebagai instrumen forensik citra.

4. Pembahasan

Evaluasi kinerja algoritma *fragile watermarking* berbasis MD5 diawali dengan analisis pada tahap pra-pemrosesan dan penyisipan. Konversi citra ke ruang *grayscale* 8-bit berkanal tunggal terbukti menjadi pendekatan komputasional yang krusial guna mengeliminasi potensi pergeseran warna (*color shifting*) akibat manipulasi bit yang tidak simetris pada citra RGB. Pada tahap penyisipan, implementasi logika bitwise secara presisi sukses mengosongkan ruang 1-bit *Least Significant Bit* (LSB) tanpa mengganggu integritas struktur 7-bit *Most Significant Bit* (MSB). Pemotongan digest MD5 menjadi 4-bit representatif dan penyisipannya ke dalam matriks blok 2×2 piksel terbukti menghasilkan tingkat ketidaktampakan (*imperceptibility*) yang sangat tinggi. Keberhasilan ini tervalidasi secara kuantitatif melalui capaian rata-rata *Peak Signal-to-Noise Ratio* (PSNR) sebesar

51,09 dB dan *Structural Similarity Index* (SSIM) sebesar 0,9977. Nilai tersebut menegaskan bahwa modifikasi spasial tingkat mikro, yang secara matematis hanya memicu deviasi intensitas piksel maksimal ±1 tingkat, mampu menyembunyikan data *watermark* tanpa menimbulkan degradasi struktural maupun derau visual yang dapat dideteksi oleh penglihatan manusia.

Pada tahap simulasi serangan, algoritma mendemonstrasikan sifat rapuh (*fragile*) yang sangat sensitif terhadap modifikasi spasial atau pemalsuan citra (*image forgery*). Substitusi matriks piksel menggunakan peta logika hasil thresholding memicu kerusakan struktural secara instan, di mana masuknya piksel asing secara otomatis menimpa dan menghancurkan "Hash Tersimpan" pada ruang LSB di zona target. Kerusakan ini berhasil diidentifikasi pada fase ekstraksi buta (*blind extraction*) sebagai *true positive*. Sistem secara deterministik memetakan ketidakcocokan antara "Hash Tersimpan" yang diekstrak secara langsung dengan "Hash Baru" hasil kalkulasi ulang dari sisa 7-bit MSB blok tersebut. Sebaliknya, pada area blok matriks yang otentik dan tidak tersentuh serangan, algoritma secara konsisten mendapati kecocokan hash, sehingga secara efektif mampu menekan indikasi deteksi palsu (*false positive*).

Analisis akhir pada visualisasi Peta Deteksi (*Detection Map*) membuktikan bahwa kombinasi pemrosesan hash MD5 dengan arsitektur pembagian citra menjadi unit mikro 2×2 piksel menghasilkan presisi pelokalan yang sangat tinggi. Sistem berhasil mencatatkan akurasi deteksi rata-rata sebesar 93,51%, yang mengonfirmasi kemampuan komputasi algoritma dalam mengisolasi batas tepi (*edge*) area manipulasi secara kongruen dengan masker ground truth. Presisi spasial tingkat mikro ini membuktikan keandalan sistem sebagai instrumen forensik digital yang tangguh untuk verifikasi autentikasi citra medis maupun dokumen legal.

Dibandingkan dengan skema *watermarking* berbasis kombinasi pixel value indicator dan MSB yang beroperasi pada satuan blok relatif besar sehingga hanya mampu melokalisasi manipulasi secara kasar [8], serta skema berbasis hash MD5 dan SHA-256 yang belum secara khusus dirancang untuk pelokalan hingga level piksel [11, 12], penelitian ini memposisikan diri pada celah granularitas blok yang belum banyak dieksplorasi, yaitu unit mikro 2×2 piksel yang memungkinkan pemetaan batas tepi manipulasi jauh lebih rinci. Posisi ini juga melengkapi kajian sistematis skema *fragile watermarking* berbasis *self-embedding* yang selama ini lebih banyak membahas mekanisme pemulihan (*self-recovery*) pada satuan blok besar [13–15], dengan menawarkan alternatif pada ekstrem granularitas terkecil yang secara empiris masih mempertahankan fidelitas visual tinggi (PSNR di atas 50 dB) sekaligus akurasi pelokalan di atas 90%.

Sebagai langkah penyempurnaan, penerapan algoritma hash berbasis kriptografi ringan (*lightweight cryptography*) direkomendasikan pada penelitian mendatang guna memangkas beban komputasi tanpa mengurangi sensitivitas deteksi sistem.

5. Kesimpulan

Secara rinci, pengujian kualitas visual menunjukkan tingkat ketidaktampakan (*imperceptibility*) yang sangat optimal, dibuktikan dengan nilai rata-rata *Peak Sig-nal-to-Noise Ratio* (PSNR) sebesar 51,09 dB dan *Structural Similarity Index* (SSIM) sebesar 0,9977, sementara pengujian simulasi serangan penimpaan piksel membuktikan sensi-tivitas sistem yang sangat tinggi melalui mekanisme ekstraksi buta (*blind extraction*) yang berhasil melokalisasi batas tepi manipulasi dengan tingkat akurasi mencapai 93,51%. Sebagai langkah pengembangan pada penelitian selanjutnya, direkomendasikan ek-spansi ruang warna (*color space*) agar algoritma dapat diimplementasikan langsung pada citra berwarna (RGB 24-bit) dengan strategi mitigasi pergeseran warna (*color shifting*), serta penerapan algoritma hash berbasis kriptografi ringan (*lightweight cryptography*) agar sistem dapat diimplementasikan untuk pemindaian forensik secara *real-time* tanpa mengurangi

sensitivitas deteksi sistem.

Pernyataan Penggunaan *Artificial Intelligence*. Penulis menggunakan ChatGPT sebagai alat bantu referensi dalam memahami dan mengeksplorasi informasi yang berkaitan dengan penelitian. Seluruh informasi yang diperoleh telah diperiksa dan diverifikasi oleh penulis menggunakan sumber yang relevan. Penulis bertanggung jawab sepenuhnya atas keakuratan, orisinalitas, dan integritas isi artikel.

Pernyataan Kontribusi Penulis. **RBW (Raihan Budi Waluyo)** berkontribusi dalam konseptualisasi, metodologi, pengumpulan dan pengolahan data, analisis penelitian, serta penyusunan naskah. **ESW (Ermadi Satria Wijaya)** berkontribusi dalam pembimbingan penelitian, pemberian arahan metodologis, serta peninjauan dan penyempurnaan naskah. **SS (Sigit Sugiyanto)** berkontribusi dalam peninjauan dan evaluasi penelitian serta memberikan masukan terhadap penyempurnaan naskah. **H (Harjono)** berkontribusi dalam peninjauan dan evaluasi penelitian serta memberikan masukan terhadap penyempurnaan naskah. Seluruh penulis telah membaca, meninjau, dan menyetujui versi akhir artikel serta bertanggung jawab atas seluruh aspek penelitian dan publikasi.

Pernyataan Konflik Kepentingan. Penulis menyatakan bahwa tidak terdapat konflik kepentingan yang berkaitan dengan penelitian, kepengarangan, dan publikasi artikel ini.

Pernyataan Pendanaan. Penelitian ini tidak menerima pendanaan khusus dari lembaga pemerintah, komersial, maupun organisasi nirlaba.

Daftar Pustaka

[1] R. Toyib, I. Saraswati, N. Faizin, and U. M. Bengkulu, "Implementasi algoritma template matching dalam mendeteksi keaslian foto pada media sosial," *JTIS (Journal Technopreneurship Information System)*, vol. 3, no. 3, pp. 32–40, 2021. https://doi.org/10.36085/jtis.v3i3.844.

[2] R. Umar, A. Fadlil, and A. I. Putra, "Analisis forensics untuk mendeteksi pemalsuan video," *J-SAKTI (Jurnal Sains Komputer dan Informasi)*, vol. 3, no. 2, pp. 193–200, 2019. https://dx.doi.org/10.30645/j-sakti.v3i2.140.

[3] D. A. Faroek and R. Umar, "Deteksi keaslian citra menggunakan metode error level analysis (ela) dan principal component analysis (pca)," *Jurnal Ilmiah Teknik Informatika*, vol. 8, no. 2, pp. 132–139, 2020. https://doi.org/10.22441/format.2019.v8.i2.006.

[4] K. N. Isnaini, H. Ashari, and A. P. Kuncoro, "Analisis forensik untuk mendeteksi keaslian citra digital menggunakan metode nist," *J. Resist. (Rekayasa Sistem Komputer)*, vol. 3, pp. 72–81, Nov. 2020. https://doi.org/10.31598/jurnalresistor.v3i2.634.

[5] A. Fauzan, "Analisis digital forensik pada kasus pemalsuan dokumen menggunakan metode national institute of standards and technology (nist)," skripsi, Universitas Amikom Purwokerto, 2023. https://eprints.amikompurwokerto.ac.id/id/eprint/1703/.

[6] A. Y. Wijaya, "Pengembangan metode block matching untuk deteksi copy-move pada pemalsuan citra," *JUTI Jurnal Ilmiah Teknologi Informasi*, vol. 15, no. 1, pp. 84–94, 2017. https://doi.org/10.12962/j24068535.v15i1.a638.

[7] Zulfan, F. Arnia, and R. Muharar, "Deteksi pemalsuan citra dengan teknik copy-move menggunakan metode ordinal measure dari koefisien discrete cosine transform," *Jurnal Nasional Teknik Elektro*, vol. 5, no. 2, pp. 165–174, 2016. https://doi.org/10.25077/jnte.v5n2.230.2016.

[8] P. Lumbanraja, "Watermarking pada citra digital menggunakan kombinasi pixel value indicator dan metode most significant bit," *METHODIKA*, vol. 5, no. 1, pp. 34–41, 2019. https://doi.org/10.46880/mtk.v5i1.416.

[9] A. R. Pambudi, Garno, and Purwantoro, "Deteksi keaslian uang kertas berdasarkan watermark dengan pengolahan citra digital," *Jurnal Informatika Polinema*, vol. 6, no. 4, pp. 69–74, 2020. https://doi.org/10.33795/jip.v6i4.407.

[10] K. Masykuroh, "Analisis pengaruh koefisien watermarking pada citra rgb dengan menggunakan perbandingan dct dan dwt," *TELKA Telekomunikasi, Elektronika, Komputasi dan Kontrol*, vol. 7, no. 1, pp. 62–70, 2021. https://doi.org/10.15575/telka.v7n1.62-70.

[11] H. Habiburrahman, "Md5 di era post-kriptografi: Studi efektivitas untuk perlindungan arsip elektronik," *Jurnal Ilmu Informasi, Perpustakaan dan Kearsipan*, vol. 14, no. 1, pp. 9–18, 2025. https://doi.org/10.24036/jiipk.v14i1.133546.

[12] I. Saputra and S. D. Nasution, "Analisa algoritma sha-256 untuk mendeteksi orisinalitas citra digital," *Prosiding Seminar Nasional Riset Informasi Science*, vol. 1, pp. 164–178, 2019. https://doi.org/10.30645/senaris.v1i0.20.

[13] L. Rakhmawati, W. Wirawan, and S. Suwadi, "A recent survey of self-embedding fragile watermarking scheme for image authentication with recovery capability," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 61, 2019. https://doi.org/10.1186/s13640-019-0462-3.

[14] C. Lin, T. Lee, Y. Chang, P. Shiu, and B. Zhang, "Fragile watermarking for tamper localization and self-recovery based on ambtc and vq," *Electronics*, vol. 12, no. 2, p. 415, 2023. https://doi.org/10.3390/electronics12020415.

[15] G. D. Su, C. C. Chang, and C. C. Lin, "Effective self-recovery and tampering localization fragile watermarking for medical images," *IEEE Access*, vol. 8, 2020. https://doi.org/10.1109/ACCESS.2020.3019832.

[16] H. Bisri and M. I. Marzuki, "Forensik citra digital menggunakan metode error level analysis, clone detection dan exif untuk deteksi keaslian gambar," *G-Tech Jurnal Teknologi Terapan*, vol. 7, no. 2, pp. 586–595, 2023. https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/2363.

[17] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic, "Comofod — new database for copy-move forgery detection," in *Proceedings of ELMAR-2013*, pp. 49–54, 2013. https://ieeexplore.ieee.org/document/6658316.