IMPLEMENTASI TEKNIK STEGANALISIS MENGGUNAKAN METODE IMPROVEMENT DIFFERENCE IMAGE HISTOGRAM PADA STEGANOGRAFI LSB

Friski Gatra Pamungkas [1)], Bambang Hidayat [2)], Nur Andini [3)]

[1),2),3)]Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom
Email : friski.gatra@gmail.com

Abstrak. *Steganografi adalah praktik penyembunyian informasi dengan cara meyisipkan pesan rahasia kedalam media lain atau dikenal sebagai file carrier seperti citra digital, audio, video ataupun IP header. Pengawasan steganografi sulit untuk dilakukan sehingga dapat disalahgunakan oleh pihak kriminal dalam menyembunyikan jejaknya, pernah ada dugaan bahwa teroris menerapkan steganografi dalam berkomunikasi. Oleh sebab itu, perlu adanya sistem yang dapat mengatasi hal tersebut. Teknik mendeteksi keberadaan pesan rahasia pada sebuah file dikenal sebagai steganalisis. Metode steganalisis berbeda untuk setiap file carrier yang digunakan. IDIH merupakan salah satu metode steganalisis yang digunakan untuk mengidentifikasi steganografi dengan file carrier berupa citra digital. IDIH merupakan improvisasi dari metode sebelumnya yaitu DIH yang pertama kali diperkenalkan oleh T. Zhang et al. Pada metode DIH, histogram LSB plane sebelum dan setelah diubah menjadi 0 dibuat dari file citra yang diindikasi memiliki pesan rahasia untuk mendapatkan rasio embedding pesan. Kemudian ditambahkan skema baru metode IDIH dalam mencari nilai error untuk menghitung rasio embedding pesan termodifikasi yang menghasilkan estimasi panjang pesan rahasia yang lebih akurat. Pada makalah ini, hasil yang dicapai adalah performa sistem yang dilihat dari tingkat akurasinya berdasarkan nilai estimasi berupa persentase panjang pesan dari citra digital yang diindikasi sebagai sebuah stego-file.*

***Kata kunci**: Steganography, Steganalysis, Histogram, Steganalysis IDIH*

1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi yang sangat pesat, menjadikan internet sebagai bagian yang tidak dapat dipisahkan dari kehidupan manusia. Karena kemudahan dan kebebasannya, internet sering kali dieksploitasi oleh pihak yang tidak bertanggung jawab dalam melakukan tindakan *cyber crime*. Maka dari itu teknik pengamanan data terus dikembangkan guna menjamin informasi yang ditransmisikan. Salah satunya yaitu steganografi.

Steganografi adalah teknik mengamankan informasi dengan prinsipnya yaitu pesan rahasia disisipkan kedalam media digital lain, sehingga tidak ada pihak yang mengetahui keberadaan pesan tersebut selain pengirim dan penerima. Media yang digunakan dikenal sebagai *carrier* atau *cover* yang dapat berupa fail citra, video, audio, teks, IP header atau media lainnya. Hingga saat ini belum ada teknik steganografi yang dapat menyembunyikan keberadaan pesan secara sempurna. Steganalisis mengacu pada suatu metode mengidentifikasi keberadaan pesan rahasia pada sebuah fail. Pengaplikasian steganalisis dalam kehidupan nyata yaitu *cyber warfare*, komputer forensik ataupun pelacakan tindakan kriminalitas di internet guna keperluan investigasi. Selain itu, steganalisis dapat dijadikan sebagai tolak ukur untuk mencari kelemahan, mengevaluasi dan mengembangkan algoritme steganografi guna meningkatkan skema penyisipan pesan yang lebih aman.

Salah satu metode steganalisis pada citra digital yaitu IDIH [3]. *Improvement Difference Image Histogram* (IDIH) merupakan pengembangan dari metode DIH yang diperkenalkan pertama kali oleh T. Zhang et al [1]. IDIH menghasilkan deteksi yang lebih akurat dibandingkan DIH, mengurangi terjadinya *false alarm* saat deteksi citra natural. Seperti pada DIH, metode ini dilakukan dengan pendekatan statistika dari karakteristik citra digital dengan melihat perbedaan histogram dari LSB *plane* citra sebelum dan setelah diisi dengan bit 0. Perbedaannya adalah ditambahkannya skema pencarian *intial* bias atau nilai eror dalam menentukan estimasi panjang pesan. Sehingga hasil akhirnya yaitu estimasi panjang pesan termodifikasi yang memiliki hasil yang lebih akurat dibandingkan metode DIH ketika pesan yang disisipkan memiliki panjang mendekati 0.

1.1. Steganografi Citra

Steganografi citra yaitu steganografi yang menggunakan media citra sebagai *cover-file* dalam proses penyisipan pesan rahasia untuk menghasilkan *stego-image. S*kema *embedding* pesan tidak sama untuk setiap *cover-file* yang digunakan, karena mengacu pada sifat dari *cover-file* yang berbeda-beda. Sebuah citra digital dipilih sebagai *cover-file* karena ukurannya yang relatif kecil, sehingga baik digunakan untuk menyisipkan sebuah pesan rahasia berupa teks.

Embedding adalah skema penyisipan pesan rahasia ke *dalam cover-file*. Pada *LSB embedding*, setiap karakter pesan rahasia dikonversi menjadi 8 bit biner berdasarkan standar ASCII. Kemudian setiap bit pesan di-*embedding* pada LSB tiap *byte* dari *cover-image*. *Embedding* pesan tidak menimbulkan perubahan yang signifikan sehingga tidak terlihat dengan jelas perbedaannya oleh mata manusia.

Karena kesederhanaan dan kemudahannya, metode LSB *embedding* baik digunakan untuk steganografi citra. metode ini dibagi menjadi beberapa teknik yaitu :

1. LSB
   Least Significant Bit (LSB) merupakan metode *embedding* pesan yang umum digunakan. Pada metode ini, bit pesan disisipkan secara berurutan dari *byte* pertama hingga *byte* terakhir cover. Teknik ini baik digunakan pada citra *grayscale* (8 bit) [5].
2. RLSB
   Randomize Least Significant Bit (RLSB) merupakan metode *embedding* pesan dengan penyisipan bit pesan dilakukan tidak secara berurutan, namun disebar secara acak.

1.2. *Difference Image Histogram* [1]

Zhang et al [1] memperkenalkan metode *difference image histogram* yang digunakan untuk mencari korelasi antara LSB planes pada setiap piksel citra digital. Metode ini melihat sifat yang dimiliki oleh citra digital sehingga dapat digunakan untuk membuat sistem steganalisis berbasis histogram yang dapat *membedakan stego-image* dengan citra natural. *Difference image histogram* [5] dijadikan sebagai alat analisa statistika.

1.3. *Improvement Difference Image Histogram* [2]

Estimasi *embedding* pesan pada citra dengan metode DIH merupakan akar dari p yang didapatkan dari metode DIH [2]. Hasil estimasi panjang pesan dari persamaan tersebut akan menimbulkan terjadinya kesalahan deteksi pada citra natural, sehingga pada metode *Improvement Difference Image Histogram* ditambahkan skema pencarian nilai eror untuk mendapatkan nilai estimasi panjang pesan termodifikasi $p_{modified}$ [2].

1.4. Sensitivitas dan Spesifisitas [3]

Pengambilan keputusan dari sistem steganalisis adalah mengklasifikasikan citra berdasarkan keberadaan pesan yang terkandung didalamnya. Dari pengklasifikasian ini, didapat *binary decision* yang mewakili kondisi ada atau tidaknya pesan pada citra dan biasanya dilakukan berdasarkan *threshold* yang telah ditentukan [3]. Dari pengujian sistem dalam mendeteksi, ada beberapa parameter yang diperhitungkan seperti sensitivitas dan spesifisitas. Sensitivitas dan spesifisitas inilah yang dapat menentukan akurasi sistem steganalisis yang dirancang.

1.5. Diagram Alir

Konfigurasi sistem untuk melakukan pendeteksian citra dijelaskan sebagai berikut:

1. Citra latih dan citra uji

Citra latih adalah citra yang dijadikan sebagai penentu nilai *threshold* yang nantinya digunakan untuk menguji sistem dengan masukan citra uji kemudian performa sistem dianalisa berdasarkan akurasinya membedakan citra natural dan *stego-image*.

2. Prepocessing

Pada tahap ini persiapan data seperti citra digital dalam format *grayscale* dan RGB, pesan yang disisipkan dan penyisipan pesan untuk membuat *stego-image*. Metode *embedding* yang digunakan adalah LSB dan RLSB.

3. Ekstraksi ciri

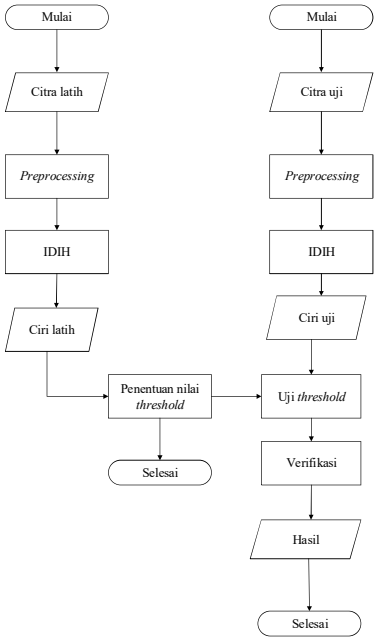
Ekstraksi ciri dilakukan untuk mengetahui ciri dari *stego-image* dan citra natural. Hasil dari ekstraksi ciri adalah nilai estimasi panjang pesan metode DIH dan IDIH yang dilakukan dengan analisa statistika.

4. Klasifikasi

Proses klasifikasi menggunakan nilai *threshold* yang didapat ketika melakukan ekstraksi ciri. Nilai rata-rata estimasi panjang citra natural dari ekstraksi ciri citra latih dijadikan *threshold* untuk diujikan pada citra uji sebagai pembanding sistem deteksi dalam membedakan *stego-image* dan citra natural.

5. Verifikasi

Verifikasi dilakukan untuk menguji akurasi sistem yang ditunjukan dari sensitivitas dan spesifisitasnya ketika melakukan klasifikasi *stego-image* dan citra natural.



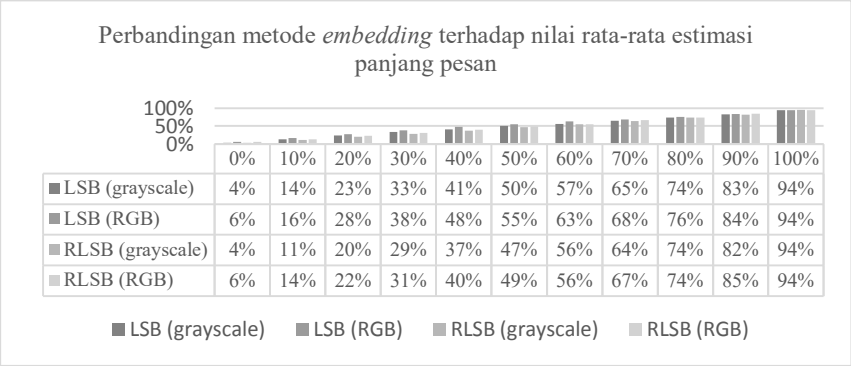
Gambar 13. Diagram alir sistem klasifikasi steganalisis histogram

2. Pembahasan

Steganografi yang diujikan terdiri dari citra latih dan citra uji. Citra latih terdiri dari 30 buah citra digital dengan resolusi 256x256px, 512x512px dan 1024x1024px dengan format *.PNG sedangkan citra uji merupakan database citra digital selain citra latih dengan jumlah 1000 buah dengan format fail yang sama namun dengan resolusi citra yang beragam. Pesan yang disisipkan adalah teks dengan format *.TXT. Citra latih digunakan untuk menganalisa sistem steganalisis dengan beberapa skenario guna menentukan nilai *threshold* pada saat melakukan pengujian pada citra uji. Pengujian citra uji djadikan sebagai tolak ukur dalam melihat performa sistem yang ditunjukan dari akurasi sistem steganalisis dalam menggolongkan citra natural dan *stego-image*.

2.1. Pengaruh Metode *Embedding* Pesan Terhadap Estimasi Panjang Pesan Menggunakan DIH

Jika dilihat berdasarkan format warna citra. Gambar 2 menunjukan bahwa citra *grayscale* menghasilkan estimasi yang cenderung lebih mendekati nilai aslinya apabila dibandingkan citra RGB dengan metode *embedding* pesan yang sama. Hal ini disebabkan karena citra RGB terdapat 3 layer yaitu merah, hijau dan biru. Dari ketiga layer tersebut distribusi *nilai difference image* tidak semuanya sama mengikuti distribusi Gaussian.



	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
LSB (grayscale)	4%	14%	23%	33%	41%	50%	57%	65%	74%	83%	94%
LSB (RGB)	6%	16%	28%	38%	48%	55%	63%	68%	76%	84%	94%
RLSB (grayscale)	4%	11%	20%	29%	37%	47%	56%	64%	74%	82%	94%
RLSB (RGB)	6%	14%	22%	31%	40%	49%	56%	67%	74%	85%	94%

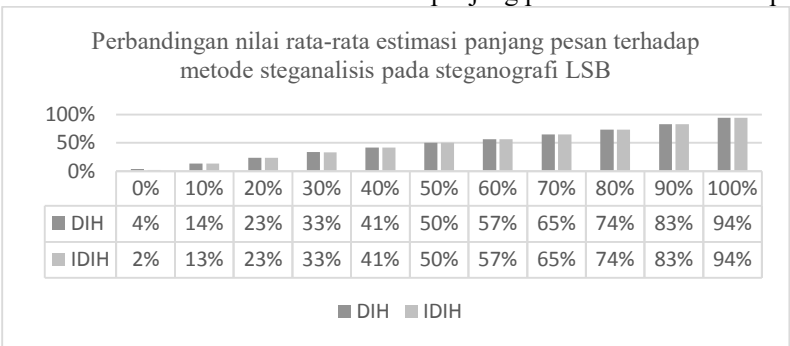
Gambar 2. Perbandingan metode *embedding* terhadap nilai rata-rata estimasi panjang pesan

Namun yang perlu diperhatikan adalah ketika sistem melakukan estimasi panjang pesan pada citra natural. Gambar 2 menunjukan hasil 4% untuk citra *grayscale* dan 6% untuk citra RGB yang seharusnya mendekati nilai aslinya yaitu 0%.

2.2. Perbandingan Estimasi Panjang Pesan Terhadap Metode DIH dan IDIH

Pada pengujian sebelumnya, steganalisis metode DIH menghasilkan nilai estimasi yang cukup menyimpang dari nilai aslinya ketika sistem diberi masukkan citra natural. Oleh sebab itu, skema untuk mencari nilai eror ditambahkan pada IDIH untuk menghasilkan nilai estimasi pesan termodifikasi.

Pada pengujian ini, citra latih dibuat dalam bentuk *grayscale* kemudian disisipi pesan menggunakan metode LSB dan RLSB dengan rasio *embedding* yang sama dengan pengujian sebelumnya. Kemudian sistem steganalisis DIH dan IDIH melakukan estimasi panjang pesan untuk dianalisa perbedaannya.



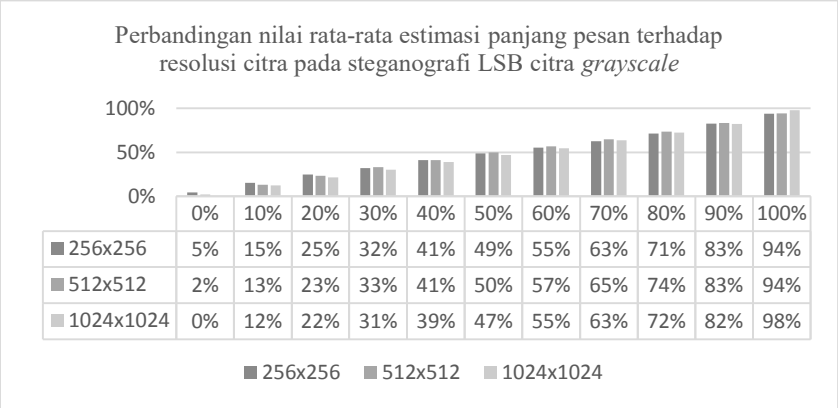
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
DIH	4%	14%	23%	33%	41%	50%	57%	65%	74%	83%	94%
IDIH	2%	13%	23%	33%	41%	50%	57%	65%	74%	83%	94%

Gambar 3. Perbandingan nilai rata-rata estimasi panjang pesan terhadap metode steganalisis pada steganografi LSB

Berdasarkan Gambar 3, nilai estimasi panjang pesan IDIH memberikan hasil yang lebih mendekati nilai aslinya dibandingkan DIH ketika citra disisipkan pesan antara 0% dan 10%. Hal ini dikarenakan pada steganalisis IDIH, pencarian nilai eror dilakukan untuk menghasilkan nilai estimasi termodifikasi.

2.3. Pengaruh Resolusi Citra Terhadap Estimasi Pesan Metode IDIH

Pengujian ini dilakukan pada citra latih dengan resolusi citra yang berbeda yaitu 256x256px, 512x512px dan 1024x1024px untuk mengetahui pengaruh resolusi pada nilai estimasi panjang pesan yang dihasilkan. Berdasarkan grafiik dari Gambar 4, citra yang diuji dengan resolusi yang berbeda untuk setiap metode *embedding* pesan yang sama menghasilkan hasil yang lebih baik ketika citra memiliki resolusi lebih tinggi, terutama ketika rasio *embedding* pesan terhadap *cover* yaitu 0%, 50% dan 100%. Hal ini dikarenakan citra yang memiliki resolusi tinggi memiliki distribusi LSB *plane* yang sangat besar. Distribusi LSB *plane* yang besar ini mengakibatkan citra dengan resolusi tinggi mendekati distribusi Gaussian.



	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
256x256	5%	15%	25%	32%	41%	49%	55%	63%	71%	83%	94%
512x512	2%	13%	23%	33%	41%	50%	57%	65%	74%	83%	94%
1024x1024	0%	12%	22%	31%	39%	47%	55%	63%	72%	82%	98%

Gambar 4. Perbandingan nilai rata-rata estimasi panjang pesan terhadap resolusi citra pada steganografi LSB citra *grayscale*

2.4. Perbandingan Metode DIH dan IDIH Terhadap Akurasi Pendeteksian *Stego-Image*

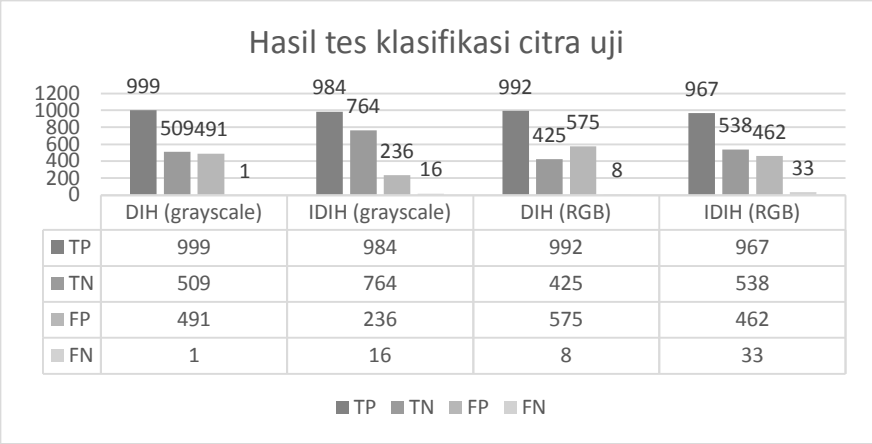
Pada pengujian ini, dilakukan pendeteksian keberadaan pesan rahasia pada citra uji. Hasil nilai rata-rata dari estimasi panjang pesan citra natural pada saat pengujian citra latih dijadikan sebagai nilai *threshold* pada sistem steganalisis dalam mengambil keputusan. Steganalisis DIH dan IDIH menggunakan *threshold* dari nilai rata-rata estimasi panjang pesan termodifikasi atau hasil estimasi dari steganalisis IDIH karena nilai inilah yang paling mendekati 0%. Dari nilai *threshold* ini diukur sensitivitas dan spesifisitasnya. Keputusan yang diambil adalah ketika citra digital yang dideteksi sistem memiliki panjang pesan diatas nilai *threshold* maka citra ini digolongkan sebagai *stego-image* selain itu dianggap sebagai citra natural.

Tabel 1. *Threshold* citra *grayscale* dan RGB

	***threshold* (*grayscale*)**	***threshold* (RGB)**		
		Biru	**Hijau**	**Merah**
256x256	0,0455	0,0369	0,0477	0,057
512x512	0,0216	0,0483	0,0431	0,0368
1024x1024	0,0033	0,0122	0,008	0,007
Rata-rata per layer	0,0235	0,0325	0.0329	0,0336
Rata-rata semua layer	0,0235	0,033		

Terlihat bahwa terdapat 2 *threshold* pada Tabel 1 yaitu *grayscale* dan RGB. Hal ini dilakukan karena sifat dari citra *grayscale* dan RGB yang berbeda, seperti yang ditunjukan pada Gambar 3. Sehingga *threshold* yang digunakan pun berbeda dengan 2,35% untuk citra *grayscale* dan 3,3% untuk citra RGB.

1000 citra uji yang ada di *database* disiapkan dalam bentuk *grayscale* dan RGB kemudian setiap citra diduplikasi. Citra yang diduplikasi ditandai sebagai citra natural dan citra hasil duplikasi ditandai sebagai *stego-image*. Citra yang ditandai sebagai *stego-image* disisipi pesan rahasia dengan rasio *embedding* 10% dengan metode *embedding* LSB dan RLSB. Sehingga secara keseluruhan terdapat 4000 buah citra yang diujikan.



Gambar 5. Hasil tes klasifikasi citra uji

Dari data yang ditunjukan pada Gambar 5, kalkulasi dilakukan untuk mendapatkan nilai sensitivitas, spesifisitas dan akurasi sistem. Hasilnya ditunjukan pada Tabel 2. Hasil yang ditunjukkan dari Tabel 2, dapat dilihat bahwa sensitivitas DIH lebih rendah dibandingkan IDIH. Hal ini dikarenakan hasil perhitungan estimasi panjang pesan DIH tidak sebaik IDIH seperti yang ditunjukan pada Gambar 3. Sehingga dapat dikatakkan bahwa sistem IDIH memiliki kemungkinan untuk mendeteksi benar sebuah *stego-image* lebih tinggi dibandingkan dengan DIH. Namun pada sistem IDIH nilai spesifisitasnya sedikit rendah dibandingkan DIH. Sehingga kemungkinan sistem melakukan kesalahan deteksi ketika diberi masukan citra natural lebih tinggi.

Tabel 8**.** Sensitivitas, spesifisitas dan akurasi

	Citra *grayscale*		Citra RGB	
	DIH	IDIH	DIH	IDIH
TPR	99,9%	98,4%	99,2%	96,7%
SPC	50,9%	76,4%	42,5%	53,8%
ACC	75,4%	87,4%	70,9%	75,3%

Secara keseluruhan sistem IDIH memiliki keakuratan yang lebih baik dibandingkan dengan DIH untuk kedua citra dengan format warna yang berbeda. Akurasi DIH dan IDIH pada citra *grayscale* adalah 75,4% dan 87,4% serta pada citra RGB adalah 70,9% dan 75,3%. Jika dilihat berdasarkan format warna citra, hasil deteksi citra RGB lebih rendah dibandingkan citra *grayscale*. Hal ini dikarenakan sistem DIH maupun IDIH kurang baik dalam melakukan estimasi panjang pesan pada citra tersebut.

3 Simpulan

Berdasarkan analisis pada pengujian sistem didapatkan beberapa kesimpulain yaitu sebagai berikut :

1. Panjang pesan yang disisipkan pada citra digital dapat memengaruhi nilai estimasi panjang pesan. Hal ini disebabkan karena untuk citra yang sama, penyisipan pesan menyebabkan nilai *difference image* LSB *plane* menjadi berbeda.
2. Metode *embedding* pesan yang digunakan memengaruhi hasil estimasi panjang pesan. Estimasi panjang pesan RLSB menghasilkan hasil yang lebih mendekati panjang pesan rahasia, karena penyisipan pesan dilakukan secara acak. Penyisipan yang dilakukan secara acak ini menyebabkan LSB *plane* citra menjadi lebih independen dibandingkan penyisipan yang dilakukan secara berurutan dengan rasio *embedding* sama.
3. Format warna citra pada *cover* yang digunakan untuk menyisipkan pesan rahasia memengaruhi hasil estimasi panjang pesan dengan kecenderungan bahwa estimasi citra *grayscale* sedikit lebih mendekati panjang pesan yang disisipkan. Hal ini disebabkan karena citra RGB memiliki 3 *layer*, dan dari ketiga layer tersebut tidak semuanya memiliki histogram dengan *difference image* mengikuti distribusi Gaussian.
4. Dimensi atau resolusi dari citra yang dipilih sebagai *cover* memengaruhi hasil estimasi panjang pesan dengan kecenderungan bahwa citra dengan resolusi lebih tinggi menghasilkan hasil yang

lebih baik. Hal ini dikarenakan citra yang memiliki resolusi tinggi memiliki distribusi LSB *plane* yang sangat besar. Distribusi LSB *plane* yang besar ini mengakibatkan citra dengan resolusi tinggi mendekati distribusi Gaussian.

5. Steganalisis metode IDIH menghasilkan estimasi panjang pesan yang lebih baik dibandingkan DIH ketika pesan yang disisipkan mendekati nol. Hal ini disebabkan pada IDIH ditambahkan skema pencarian nilai eror dalam menentukan estimasi panjang pesan.
6. Steganalisis metode IDIH memiliki akurasi yang lebih baik dibandingkan DIH dalam membedakan *stego-image* dan citra natural. Hal ini dikarenakan kejadian *false alarm* IDIH lebih rendah dibandingkan DIH meskipun persentase sensitivitasnya sedikit dibawah DIH. Namun secara keseluruhan akurasi IDIH lebih baik untuk citra *grayscale* maupun RGB.

Daftar Pustaka

[1] T. Zhang and X. Ping "*A New Approach to Reliable Detection of LSB Steganography in Natural Images*", *IEEE Signal Processing*, Vol.83, pp. 2085-2093, 2003.

[2] Ciprian Iulian and Valentin Garban. *"Research On Estimation Length Of Hidden Message"*, International DAAAM Symposium, Vol. 21, No. 1, 2010.

[3] Ashraf M. Emam and Mahmoud M. Ouf "*Performance Evaluation of Different Universal Steganalysis Techniques in JPG Files*", *Annales UMCS Informatica AI XII,* 2012.

[4] Alatas, Putri. Implementasi Teknik Steganografi dengan Metode LSB pada Citra Digital, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma : Jakarta, 2009.

[5] Yambem Jina Chanu, Kh. Manglem Singh and Themrichon Tuithung. "*Image Steganography and Steganalysis: A Survei"*, International Journal of Computer Applications, vol. 52, no. 2, August 2012.

[6] Sulistyo, W. Analisis Penerapan Metode Median Filter untuk Mengurangi Noise pada Citra Digital. *Bali : Konferensi Nasional Sistem dan Informatika*, 2009.