

VIDEO STEGANOGRAFI MENGGUNAKAN METODE *ENHANCED LEAST SIGNIFICANT BIT* (ELSB) PADA *FRAME* YANG DIPILIH BERDASARKAN DETEKSI ENERGI *MEL-FREQUENCY CEPSTRAL COEFFICIENT* (MFCC)

STEGANOGRAPHY VIDEO USING ENHANCED LEAST SIGNIFICANT BIT (ELSB) METHOD IN SELECTED FRAME BASED ON ENERGY DETECTION OF MEL-FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

Rizky Setyaningrum¹, Bambang Hidayat², Nur Andini³

^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik, Universitas Telkom

¹rizkysetyaningrum@students.telkomuniversity.ac.id, ²bhidayat@telkomuniversity.ac.id

³nurandini@telkomuniversity.ac.id

Abstrak

Penyisipan informasi ke dalam file digital adalah hal yang penting, karena persebaran informasi sangat mudah dilakukan. Steganografi merupakan salah satu teknik penyembungian pesan, sehingga hanya pengirim dan penerima yang mengetahui informasi rahasia dalam file tersebut. Dalam tugas akhir ini dilakukan simulasi sistem steganografi pada video dengan menggunakan metode *Enhanced Least Significant Bit* (ELSB) yang merupakan modifikasi dari metode LSB. Penyisipan dilakukan pada citra dengan pemilihan frame menggunakan energi Mel-Frequency Cepstral Coefficient (MFCC) yaitu mengekstraksi ciri sinyal suara berdasarkan karakter respon frekuensi suara. Hasil yang diperoleh dari pengujian beberapa ukuran citra pesan dan ukuran cover video, serta melakukan pemilihan di beberapa nilai threshold, didapatkan nilai-nilai *Peak Signal to Noise Ratio* (PSNR) yang baik, yaitu masih diatas 40dB. Nilai *Mean Square Error* (MSE) terbesar yaitu 3,8. Hasil *Mean Opinion Score* (MOS) yang didapatkan dalam rentang baik. Pada saat sistem diberikan noise Gaussian baik pada gambar maupun pada audio, sistem cukup mampu untuk bertahan. Pada serangan sebesar 15dB pada gambar, BER yang didapatkan <3%.

Kata kunci : Steganografi, video, citra, ELSB, MFCC

Abstract

The insertion of information into digital files is important, because the spread of information is very easy to do. Steganography is a technique penyembungian messages, so that only the sender and recipient know the confidential information in those files. In this final task performed a simulation system of steganography in the video by using the Enhanced method of Least Significant Bit (ELSB) which is a modification of the method of the LSB. The insertion is done on the image with the selection of frames using energy Mel-Frequency Cepstral's (MFCC) i.e. extracting characteristics of sound signal based on frequency response character of sound. The results obtained from testing several image sizes and the size of the video cover, as well as conducting the election at some threshold value, the obtained values of the Peak Signal to Noise Ratio (PSNR) is good, that is still above the 40dB. The value of the Mean Square Error (MSE) i.e. 3,8. The results Mean Opinion Score (MOS) obtained in the good range. At the time of the system given Gaussian noise on either the image or audio system is quite capable to stick around. On the attack of 15dB on image, BER obtained <3%.

Key words: Steganography, video, image, ELSB, MFCC

1 Pendahuluan

Berkembangnya teknologi informasi memudahkan pertukaran informasi melalui media digital yang membawa dampak positif dan dampak *negative*. Dampak *negative* yang dapat mengganggu jalannya pertukaran informasi diantaranya adalah pencurian data yang kemudian disalahgunakan oleh pihak yang tidak bertanggungjawab. Oleh karena itu untuk menjamin keamanan dan kerahasiaan data diperlukan teknik-teknik yang berfungsi untuk mengamankan data tersebut diantaranya adalah steganografi.

Steganografi merupakan teknik yang digunakan untuk menyembunyikan informasi dalam suatu wadah atau media. Dalam proses steganografi suatu wadah atau media tersebut tidak mengalami perubahan yang berarti, dengan kata lain informasi yang telah disisipkan tidak mengubah wadah atau media secara kasat mata. Wadah atau media yang dapat digunakan dalam teknik steganografi antara lain dapat berupa gambar, data audio maupun video.

Metode yang cukup sering digunakan dalam teknik steganografi salah satunya adalah *Least Significant Bit* (LSB), dalam metode ini informasi atau pesan rahasia menggantikan bit paling rendah dalam media yang disisipi. Metode ini memang mudah diimplementasikan namun cukup mudah juga bagi seseorang untuk mengetahui letak informasi yang disembunyikan khususnya pada video.

Pada penelitian sebelumnya [4] dilakukan penyembunyian pesan pada media *host* audio dengan metode ELSB tanpa menentukan letak tempat pesan rahasia yang akan disembunyikan. Pada penelitian tersebut media yang digunakan untuk menyisipkan pesan adalah audio, pesan yang disisipkan juga merupakan audio. Sedangkan pada penelitian [5] *Mel-Frequency Cepstral Coefficient* digunakan untuk klasifikasi suara burung dengan ciri hasil dari metode MFCC tersebut.

Oleh karena itu, pada tugas akhir ini metode yang digunakan adalah modifikasi dari *Least Significant Bit* (LSB) yaitu *Enhanced Least Significant Bit (ELSB)* dengan tujuan memaksimalkan area-area bit yang akan disisipi informasi pada data video untuk menghasilkan kualitas yang lebih baik. Metode ini dapat dilakukan dengan dua cara, pertama dengan mengacak jumlah bit dari file *host* yang digunakan untuk *embedding* pesan, dan yang kedua adalah mengacak sampel *host* yang berisi pesan rahasia bit selanjutnya. Penyisipan pesan yang berupa teks ini dilakukan pada *frame* atau gambar (*image*) saat terdapat suara dalam video. Pemilihan *frame* audio menggunakan *Mel-Frequency Cepstral Coefficient* (MFCC) yaitu mengekstraksi ciri sinyal suara berdasarkan karakter respon frekuensi suara. Keluaran yang diharapkan dari tugas akhir ini adalah kualitas video yang telah disisipi steganografi tidak banyak berbeda dengan video asli serta mendapatkan pengukuran kualitas yang baik dengan parameter Waktu Komputasi, *Peak Signal to Noise Ratio* (PSNR), *Mean Opinion Score* (MOS), *Bit Error Rate* (BER), dan *Mean Square Error* (MSE).

2 Teori dan Tahap Perancangan

A. Enhanced Least Significant Bit (ELSB)[1]

Proses penyisipan pesan dilakukan dengan metode *Enhanced Least Significant Bit* (ELSB) yang merupakan modifikasi dari metode *Least Significant Bit* (LSB). Metode ini dapat dilakukan dalam dua cara, yaitu dengan mengacak jumlah bit dari file *host* yang digunakan untuk menyisipkan pesan rahasia atau dengan mengacak *sample* dari *host* yang disisipi bit pesan rahasia berikutnya. Pemilihan bit yang akan disisipi pesan rahasia dalam proses ELSB memiliki aturan sebagai berikut :

Tabel 1(a)Skema pemilihan bit yang akan disisipi[1](b) Skema pemilihan Sampel[1]

1 st MSB	2 nd MSB	Secret message bit
0	0	3 rd LSB
0	1	2 nd LSB
1	0	1 st LSB
1	1	1 st LSB

(a)

1 st MSB	2 nd MSB	3 rd MSB	Sample containing next secret message bit
0	0	0	i + 1
0	0	1	i + 2
0	1	0	i + 3
0	1	1	i + 4
1	0	0	i + 5
1	0	1	i + 6
1	1	0	i + 7
1	1	1	i + 8

(b)

B. Mel-Frequency Cepstral Coefficient (MFCC)

MFCC merupakan metode untuk melakukan feature extraction pada sinyal dengan mengekstraksi ciri suara yang merupakan adaptasi dari sistem pendengaran manusia. Metode ini menghasilkan data seminimal mungkin dan mereplikasi organ pendengaran manusia dalam melakukan persepsi terhadap sinyal suara [2]. Berikut ini merupakan tahapannya[3]:

1. Pre-emphasis

Pre-emphasis bertujuan untuk mengurangi nilai amplitude yang rendah yang kemungkinan besar merupakan *noise* dan memperbesar nilai-nilai yang konstan sehingga didapatkan nilai yang benar-benar merepresentasikan audio asli.

$$H(z) = 1 - \alpha z^{-1}, 0.9 \leq \alpha \leq 1.0 \tag{2.1}$$

2. Frame Blocking

Pada proses ini sinyal dibagi menjadi beberapa *frame* yang dipotong menjadi ukuran yang lebih kecil (dibuat menjadi *frame-frame*). Pemotongan sinyal (*framing*) dilakukan dengan membagi panjang data audio dengan jumlah *frame* video. Tujuan dari melakukan *overlapping* adalah agar tidak ada sinyal informasi yang hilang akibat pemotongan.



(2.2)

3. Windowing

Proses selanjutnya adalah proses *windowing*. Proses *windowing* ini dilakukan pada setiap *frame* untuk mengurangi kebocoran *spectral* dan mengurangi efek diskontinuitas di awal dan akhir masing-masing *frame*. *Window Hamming* digunakan karena memiliki panjang filter yang sama, *window Hamming* mampu memberikan nilai

transition bandwidth yang tidak terlalu besar. Dengan ketepatan frekuensi *cut-off* yang cukup mendekati spesifikasi dan redaman yang cukup baik sehingga *noise* yang dihasilkan tidak terlalu besar.

$$H(k) = 0.54 - 0.46 \cos\left(\frac{2\pi k}{N-1}\right), 0 \leq k \leq N-1 \tag{2.3}$$

4. *Discrete Fourier Transform (DFT)*

Pada proses ini dilakukan proses *Fast Fourier Transform (FFT)* yang merupakan realisasi dari proses DFT yaitu mentransformasikan sinyal dari domain waktu ke domain frekuensi untuk mempermudah perhitungan. Jumlah titik FFT yang digunakan adalah nilai kelipatan dua terdekat dengan jumlah sampel suatu *frame*.

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N}, 0 \leq k \leq N-1 \tag{2.4}$$

5. *Mel-Frequency Warping*

Mel-frequency Warping umumnya dilakukan dengan menggunakan *Filterbank*. *Filterbank* adalah salah satu bentuk dari *filter* yang dilakukan dengan tujuan untuk mengetahui ukuran dari *band frequency* dalam sinyal suara. Tugas akhir ini menggunakan frekuensi sampling 44100 Hz dengan *range* frekuensi 0 – 22050 Hz.

Nilai-nilai skala Hz tersebut kemudian diubah ke dalam skala ‘*Mel*’, lalu rentang nilai tersebut dibagi menjadi beberapa nilai sejumlah *filterbank*, dalam hal ini jumlah *filterbank* yang digunakan adalah 20 buah. Selanjutnya nilai-nilai hasil konversi ‘*Mel*’ tersebut dikonversi kembali ke dalam skala Hz. Kemudian dibuatlah *filterbank* dengan M filter, dimana m adalah nomor triangular filter pada *filterbank* menggunakan persamaan berikut.

$$H(m, k) = \begin{cases} 0, & 0 \leq k < \left(\frac{f_m}{2}\right) \\ \frac{k - \left(\frac{f_m}{2}\right)}{\left(\frac{f_m}{2}\right) - f_{m-1}}, & \left(\frac{f_m}{2}\right) \leq k < f_m \\ \frac{f_{m+1} - k}{f_{m+1} - f_m}, & f_m \leq k < f_{m+1} \\ 0, & f_{m+1} \leq k < N \end{cases} \tag{2.5}$$

6. *Discrete Cosine Transform (DCT)*

Proses perkalian antara *Mel filterbank* dengan spectral daya periodogram hasil proses FFT menghasilkan energi dari *filterbank*. Hasil perkalian spectral daya dengan *filterbank* tersebut kemudian dihitung nilai *log* dari masing-masing energi tersebut.

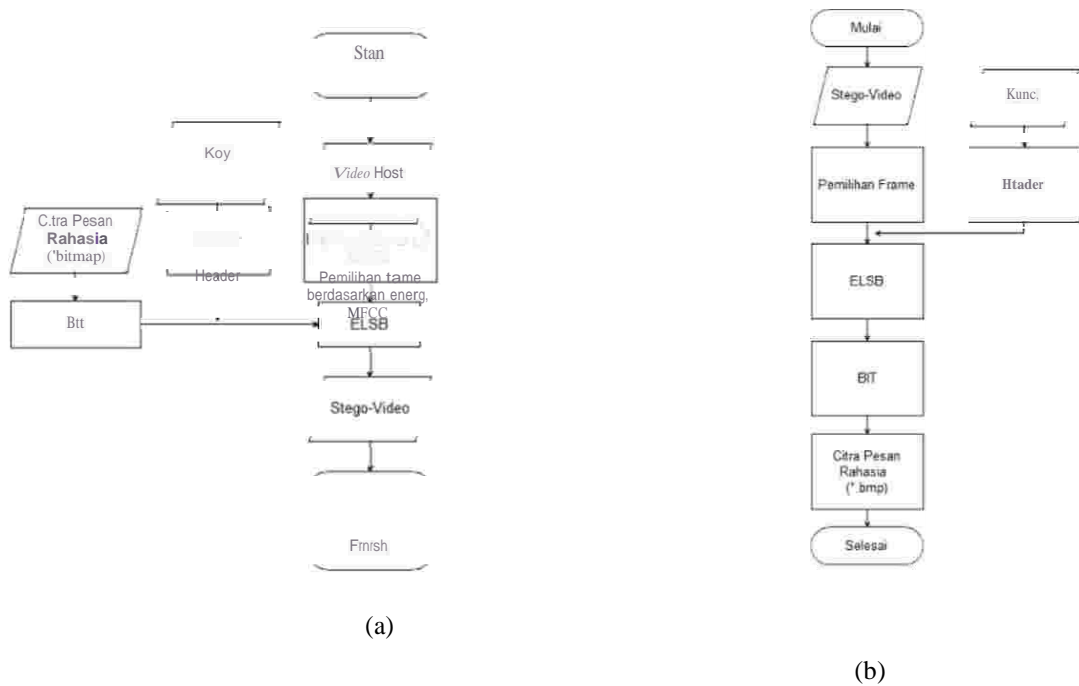
7. *Cepstral Lifting*

Proses DCT akan menghasilkan koefisien MFCC dimana koefisien MFCC yang dihasilkan akan diperhalus melalui proses *cepstral lifting* sehingga akan lebih baik digunakan saat klasifikasi nanti.

C. Perancangan Sistem

Berikut ini merupakan perancangan sistem yang digunakan pada tugas akhir ini :

1. Proses pertama yang dilakukan adalah menentukan video yang akan disisipi pesan rahasia atau disebut juga *host* dengan ketentuan ukuran maksimal yang telah ditentukan.
2. Setelah *video host* ditentukan, proses selanjutnya adalah pemilihan *frame* citra yang akan disisipi pesan dengan melihat parameter energi MFCC pada sinyal audio yang sesuai dengan *threshold* yang telah ditentukan.
3. *Embedding* atau proses penyisipan citra pesan rahasia dilakukan dengan metode ELSB (Enhanced Least Significant Bit). Pesan yang dimasukkan pada video memiliki format yang telah ditentukan yaitu *.avi. Pada video citra pesan rahasia tersebut dimasukkan ke dalam *frame YCbCr*.
4. Diperoleh video yang telah mengalami proses steganografi, video telah memuat citra pesan rahasia yang akan dikirimkan dengan kunci yang terdapat pada video tersebut.
5. Ekstraksi merupakan proses pengambilan citra pesan rahasia dengan metode yang sama dengan proses *embedding* yang telah dilakukan sebelum video dikirimkan kepada penerima. Sebelum proses ekstraksi dilakukan terlebih dahulu penerima memasukkan kunci yang sesuai dengan kunci yang dimasukkan pada saat proses penyisipan. Apabila kunci yang dimasukkan pada penerima telah sesuai, kemudian dilanjutkan dengan proses MFCC dan kemudian ELSB untuk mengekstraksi citra pesan rahasia dari video.
6. Citra pesan rahasia berformat *bmp berhasil didapatkan oleh penerima.



Gambar 1(a)Diagram alir proses penyisipan (b)Diagram alir proses ekstraksi

D. Parameter pengujian

Pada tugas akhir ini dilakukan pengujian yang dilakukan pada hasil stego-video menentukan kualitas dari sistem yang dibuat. Parameter yang akan diuji diantaranya adalah

- a. *Peak Signal to Noise Ratio* (PSNR) dan *Mean Square Error* (MSE)

PSNR merupakan nilai perbandingan antara harga maksimum dari intensitas citra terhadap error citra yaitu MSE. MSE adalah nilai yang menyatakan rata rata kuadrat error, dalam hal ini error menyatakan selisih antar citra dimana kedua citra yang dibandingkan memiliki ukuran yang sama. Oleh karena itu, sebelum dapat menghitung PSNR suatu citra kita harus menghitung nilai MSE terlebih dahulu. Untuk menghitung nilai MSE digunakan persamaan berikut ini:

$$MSE = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N (C_{xy} - C_{xy'})^2 \tag{2.6}$$

Sementara perhitungan nilai PSNR dilakukan dengan menggunakan persamaan berikut ini:

$$PSNR = 10 \log_{10} \left(\frac{C_{max}^2}{MSE} \right) \tag{2.7}$$

PSNR yang semakin besar menandakan bahwa kualitas citra semakin bagus, hal ini karena error antara kedua citra semakin kecil.

- b. *Bit Error Rate* (BER)

Merupakan persentase bit yang mengalami error dengan jumlah keseluruhan bit pada citra. Semakin kecil nilai BER, semakin bagus kualitas citra, karena semakin kecil jumlah bit yang mengalami error.

- c. Waktu Komputasi

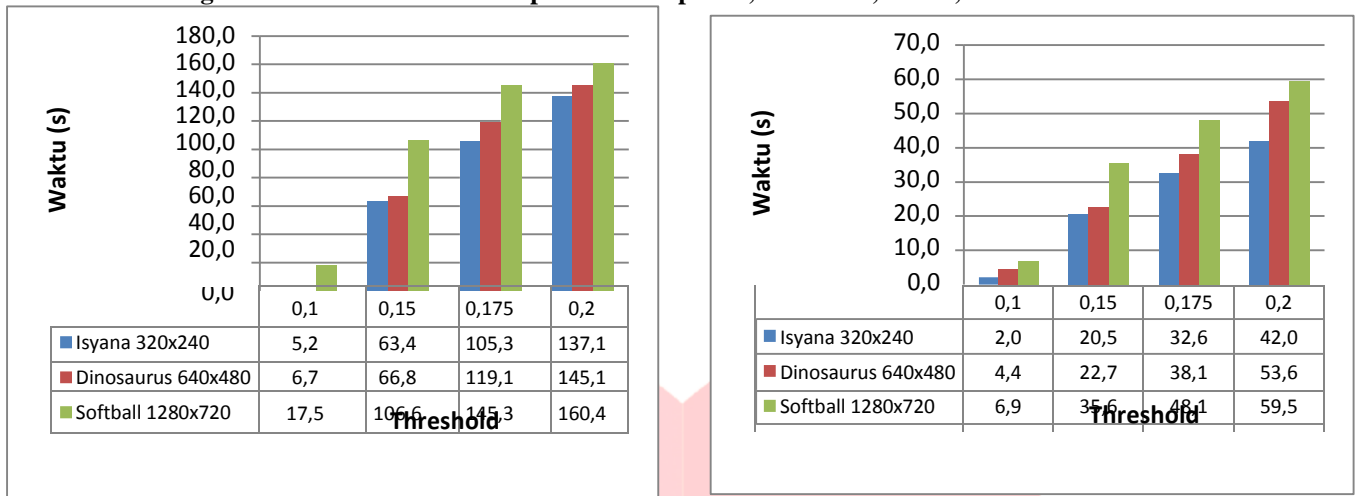
Waktu komputasi adalah waktu yang dibutuhkan sistem untuk melakukan suatu proses. Waktu komputasi sistem dihitung dari mulainya proses hingga proses tersebut selesai.

- d. *Mean Opinion Score* (MOS)

Merupakan parameter subjektif yang membandingkan perubahan objek (dalam tugas akhir ini objek yang digunakan adalah video) sebelum dan sesudah disisipi pesan rahasia. Nilai MOS akan didapatkan setelah melakukan *survey* minimal pada 30 responden.

3 Hasil dan Analisis

A. Pengaruh nilai *threshold* terhadap waktu komputasi, nilai BER, PSNR, dan MSE



(a)

(b)

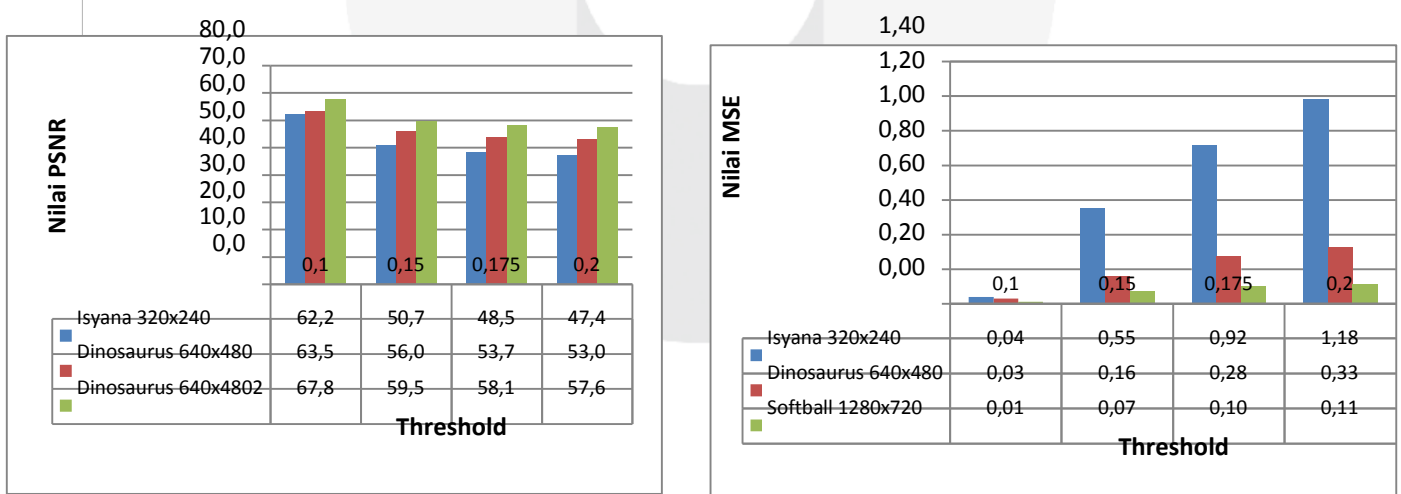
Gambar 2(a) Pengaruh nilai *threshold* terhadap waktu penyisipan (b) Pengaruh nilai *threshold* terhadap waktu ekstraksi

Dapat dilihat dari tabel diatas bahwa semakin besar *threshold* yang digunakan maka semakin banyak *frame* yang terpilih oleh proses MFCC, kemudian semakin besar waktu yang digunakan untuk proses penyisipan dan ekstraksi seperti yang dapat terlihat dari gambar 2.

Tabel 2 Pengaruh nilai *threshold* terhadap nilai BER

Threshold	Nilai BER Video		
	Isyana 320x240	Dinosaurus 640x480	Softball 1280x720
0.1	0%	0%	0%
0.15	0%	0%	0%
0.175	0%	0%	0%
0.2	0%	0%	0%

Tabel diatas merupakan tabel pengaruh nilai *threshold* terhadap nilai BER, dapat dilihat bahwa nilai BER yang didapat adalah nol. Hal ini menunjukkan bahwa pada proses penyisipan maupun proses ekstraksi tidak terdapat kesalahan bit pada citra image.



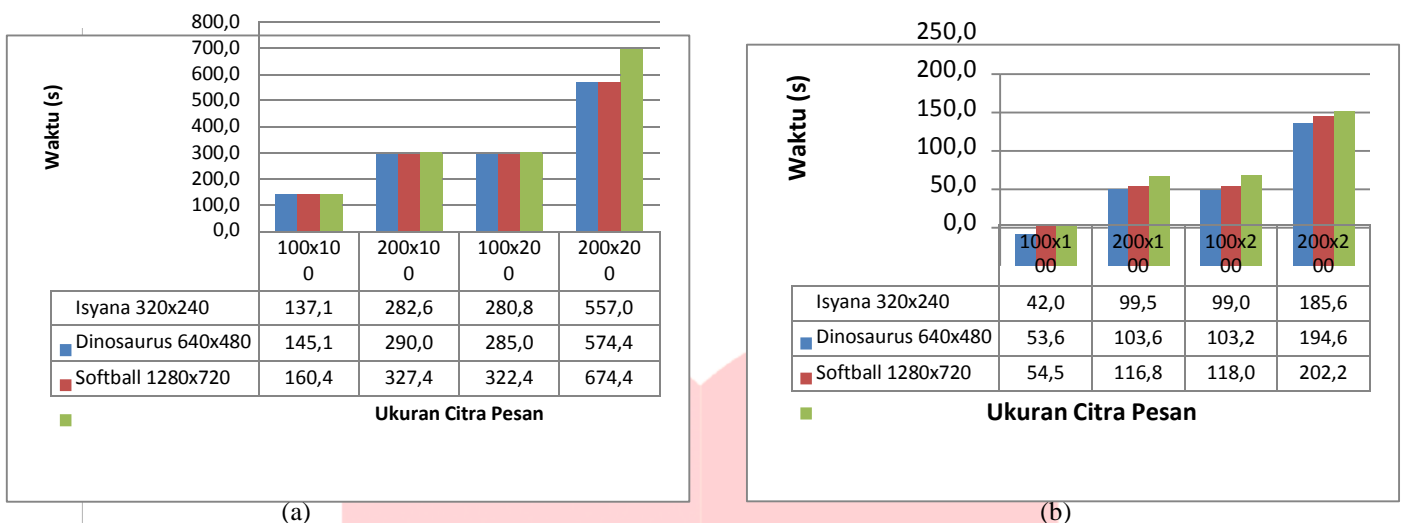
(a)

(b)

Gambar 3(a) Pengaruh nilai *threshold* terhadap nilai PSNR (b) Pengaruh nilai *threshold* terhadap nilai MSE

Dapat dilihat dari grafik diatas bahwa semakin besar *threshold* yang digunakan pada pengujian, maka nilai PSNR yang didapat semakin kecil, berbanding terbalik dengan nilai MSE yang didapatkan. Nilai MSE semakin besar karena kesalahan bit pada proses pengujian juga semakin besar.

B. Pengaruh ukuran cover video dan ukuran citra pesan terhadap waktu komputasi, nilai BER, PSNR, dan MSE



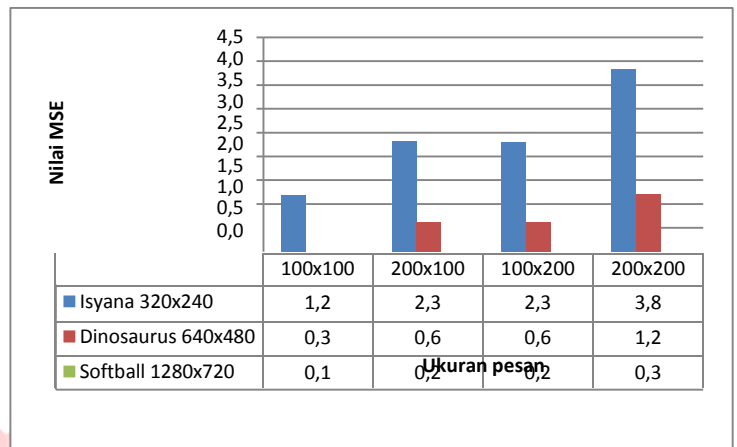
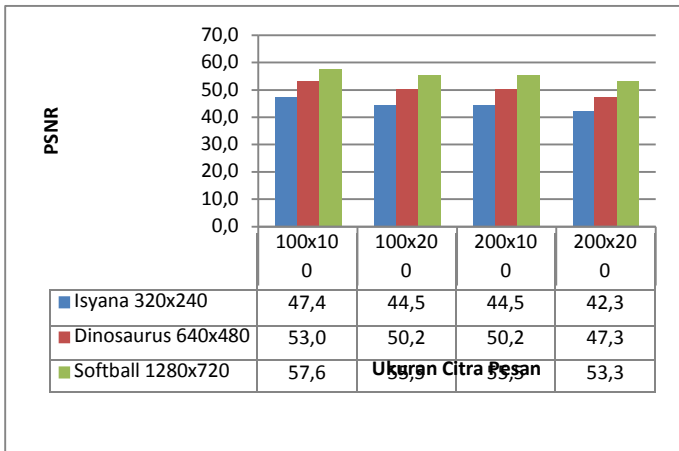
Gambar 4(a) Pengaruh ukuran *cover* dan ukuran pesan terhadap waktu penyisipan (b) Pengaruh ukuran *cover* dan ukuran pesan terhadap waktu ekstraksi

Dari hasil pengujian tersebut dapat diketahui bahwa semakin besar ukuran citra pesan yang disisipkan pada *video cover*, maka semakin lama pula waktu untuk penyisipan dan juga ekstraksi pesan. Hal ini dikarenakan semakin besar ukuran citra pesan rahasia maka semakin banyak pesan yang akan disisipkan sehingga sistem membutuhkan waktu yang lebih lama untuk proses steganografinya. Semakin besar video yang dipakai sebagai *host*, maka semakin besar pula waktu untuk proses steganografinya baik pada penyisipan maupun pembacaan citra pesan. Karena semakin besar ukuran video, maka semakin banyak pesan yang dapat disisipkan pada video tersebut karena kapasitasnya yang besar.

Tabel 3 Pengaruh ukuran *cover* dan ukuran pesan terhadap nilai BER

Sisipan	BER Video		
	Isyana 320x240	Dinosaurus 640x480	Softball 1280x720
100x100	0%	0%	0%
200x100	0%	0%	0%
100x200	0%	0%	0%
200x200	0%	0%	0%

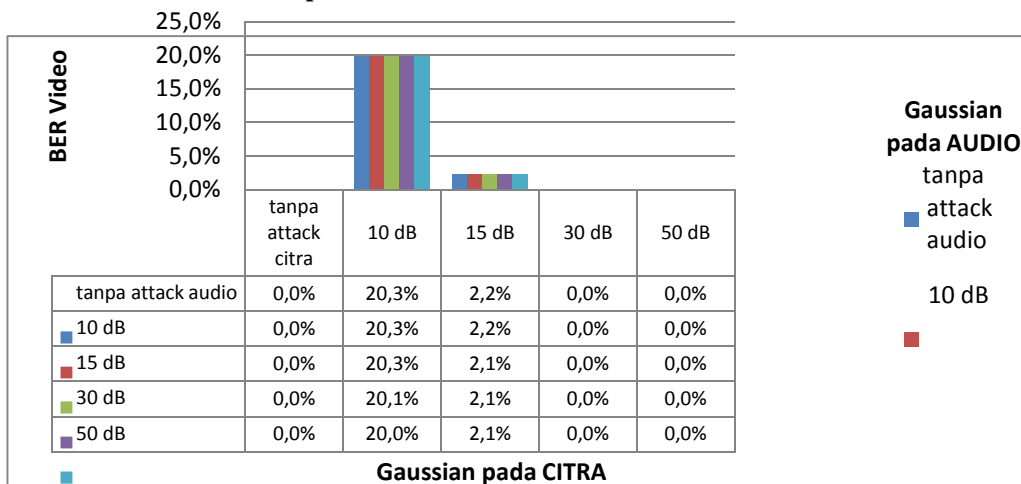
Dapat dilihat pada tabel diatas, ukuran video dan ukuran citra pesan yang disisipkan tidak mempengaruhi nilai BER , karena ketika proses steganografi pesan akan disisipkan ke semua *frame* yang terpilih dengan proses MFCC, dan ketika proses ekstraksi dilakukan pesan yang dibaca dari semua *frame* akan dirata-rata untuk meminimalisasi kesalahan bit sehingga dapat diperoleh citra pesan yang sesuai dengan yang disisipkan pada pengirim dan memiliki nilai BER yang rendah mendekati nol.



(a) Pengaruh ukuran cover dan ukuran citra pesan terhadap nilai PSNR (b) Pengaruh ukuran cover dan ukuran citra pesan terhadap nilai MSE

Dapat dilihat dari grafik diatas bahwa semakin besar ukuran video yang digunakan sebagai tempat penyisipan maka semakin besar pula nilai PSNR, dan demakin kecil nilai MSE maka tingkat kemiripan video cover dengan stego video semakin besar karena kesalahan bit pada stego video semakin kecil.

C. Ketahanan sistem terhadap noise



Gambar 6 Pengaruh noise AWGN pada citra dan audio video terhadap nilai BER

Dapat dilihat bahwa nilai BER setelah video diberi noise AWGN sebesar 10dB dan 15dB pada citra berubah. Noise yang diberikan secara acak pada citra dan audio ini mempengaruhi nilai BER pada saat nilai noise-nya kecil (di bawah 30dB). Semakin kecil ukuran noise yang diberikan maka bit yang error pada hasilnya akan semakin besar sehingga nilai BER semakin besar.

D. Pengujian MOS

Tabel 4 Hasil survey MOS

	Video A	Video B	Video C	Video D
Nilai Rata-Rata MOS	4.2	3.8	3	3.0

Setelah melakukan survey terhadap 30 responden didapatkan nilai rata-rata untuk video dengan sisipan 100x100 adalah 3.85, untuk video dengan sisipan 200x200 nilai rata-rata yang didapatkan adalah 3.4, untuk video dengan noise pada gambar 10dB dan pada audio 10dB adalah 3.2, sedangkan untuk video dengan noise pada gambar sebesar 15dB dan pada audio sebesar 15dB adalah 3,01. Berdasarkan nilai yang didapatkan dapat disimpulkan bahwa kualitas stego-video adalah baik.

4 Kesimpulan

Sistem steganografi yang dibuat dapat bekerja dengan baik, dengan parameter yang dihasilkan antara lain BER pada proses steganografi 0%, yang berarti tidak ada kesalahan dalam proses penyisipan dan ekstraksi. PSNR>30dB yang menunjukkan bahwa sistem memiliki hasil yang cukup maksimal, dan juga dengan MSE<0.5 yang menunjukkan bahwa tingkat kesalahan bit pada proses cukup kecil sehingga sistem dapat dikatakan baik juga dengan hasil MOS yang mencapai 3. Pada saat diberi noise, sistem cukup memiliki ketahanan yang baik walau ketika nilai tertentu tingkat akurasi menurun.

Daftar Pustaka :

- [1] Muhammad Asad, Junaid Gilani, Adnan Khalid. 2011. *An Enhanced Least Significant Bit Modification Technique for Audio Steganography*, Telecommunication Engineering Department, University of Engineering and Technology Taxila: Pakistan
- [2] Dadang Gunawan dan Juwono, Filbert Hilman, *Pengolahan Sinyal Digital Dengan Pemrograman Matlab*, Graha Ilmu, Yogyakarta, 2012.
- [3] Setiawan, Angga. 2011. "Aplikasi Pengenalan Ucapan Dengan Ekstraksi Mel-Frequency Cepstrum Coefficient (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) Untuk Mengoperasikan Kursor Computer,". Semarang : Jurusan Teknik Elektro Fakultas Teknik Universitas Diponegoro
- [4] Hartoko, Carolus Ferdy Setiaji, 2014. "Analisis dan Simulasi Steganografi pada Sinyal Audio Tiga Dimensi Berbasis Enhanced Least Significant Bit". Fakultas Teknik Elektro, Universitas Telkom
- [5] Hanesia, Wulandary Ika, 2015. "Klasifikasi Suara Lovebird dengan Metode Mel-Frequency Cepstral Coefficient (MFCC) dan Fuzzy Logic". Fakultas Teknik Elektro. Universitas Telkom