

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

Video Steganography Using Enhanced Least Significant Bit (ELSB) Method In Selected Frame Based On Silence Detection Mel-Frequency Cepstral Coefficient (MFCC)

Garizah Ganih Pranoto¹, Bambang Hidayat², Nur Andini³

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

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

³nurandini@telkomuniversity.ac.id

Abstrak

Pada tugas akhir ini dilakukan simulasi sistem steganografi pada video dengan menggunakan metode *Enhanced Least Significant Bit (ELSB)* yaitu modifikasi dari metode *Least Significant Bit (LSB)*. Penyisipan dilakukan pada gambar (*image*) saat tidak adanya suara (*silence*) dalam video dengan menggunakan *Mel-Frequency Cepstral Coefficient (MFCC)* yaitu mengekstraksi ciri sinyal suara berdasarkan karakter respon frekuensi suara.

Hasil yang diperoleh adalah sistem steganografi dengan waktu komputasi tercepat 1,38774 detik pada saat penyisipan dan 0,1635 detik pada saat ekstraksi. Sistem juga menghasilkan nilai akurasi sebesar 100% dan nilai PSNR mencapai 73,5329 dB dengan BER dan CER sama dengan 0 saat tidak adanya serangan *noise* Gaussian. Sistem steganografi yang dibuat tahan terhadap serangan *noise* Gaussian pada citra dengan nilai mean=0 hingga variansi 1×10^{-7} . Hasil MOS yang didapatkan dari survey terhadap 30 koresponden memiliki nilai rata-rata total 4,5 yang berarti kualitas video yang tersisipi adalah baik.

Kata kunci: Steganografi, video, ELSB, MFCC

Abstract

In this final project, it has been simulated steganography system in video using *Enhanced Least Significant Bit (ELSB)* is a modification of the method of *Least Significant Bit (LSB)*. Insertion will be carried out on the image (*image*) when there is no sound (*silence*) in the video by using *Mel-Frequency Cepstral Coefficient (MFCC)* which is extracting characteristic sound signal based on the character of the sound frequency response.

The result is steganography system with minimal computing time is 1.38774 seconds when embedding and 0.1635 when extracting. The system also produces 100% accuracy and PSNR value reach 73.5329 dB with BER and CER is 0 when there is no attack from Gaussian noise. This steganography system is resistant to *noise* Gaussian in image with mean=0 until variance 1×10^{-7} . The average value of MOS result from 30 correspondent is 4.5 that means the quality of the stego-video is good.

Keywords: Steganography, video, ELSB, MFCC

1. Pendahuluan

Pesatnya perkembangan teknologi multimedia, jaringan komputer, dan jaringan *internet* membuat pengiriman pesan baik itu berupa teks, audio, citra, dan video menjadi mudah. Kemudahan tersebut memungkinkan pesan bocor ke pihak ketiga. Oleh karena itu, diperlukan suatu pengamanan agar pesan tersebut dapat terkirim dalam keadaan utuh dan tidak bocor ke pihak ketiga. Untuk menjamin keamanan dan kerahasiaan pesan digunakanlah teknik steganografi, yaitu proses penyisipan pesan pada data lainnya yang dapat berupa audio, *image* maupun video. Pada tugas akhir ini dilakukan simulasi teknik steganografi pada video dengan format .avi menggunakan metode *Enhanced Least Significant Bit (ELSB)* dengan pemilihan *frame* dilakukan di audio saat tidak adanya suara (*silence*) dengan menggunakan *Mel-Frequency Cepstral Coefficient (MFCC)* yaitu mengekstraksi ciri sinyal suara berdasarkan karakter respon frekuensi suara. Data rahasia yang disembunyikan berupa teks dengan format .txt. Performansi sistem diuji melalui perhitungan *Mean Square Error (MSE)*, *Peak Signal to Noise Ratio (PSNR)*, *Mean Opinion Score (MOS)* dan waktu komputasi.

2. Landasan Teori

A. Steganografi

Steganografi adalah seni dan ilmu menulis atau menyembunyikan pesan tersembunyi dengan suatu cara sehingga selain si pengirim dan si penerima, tidak ada seorangpun yang mengetahui atau menyadari bahwa ada suatu pesan rahasia. Kata steganografi berasal dari bahasa Yunani “*steganos*” yang artinya tersembunyi atau terselebung, dan “*graphein*” yang artinya menulis. [1]. Pada umumnya terdapat dua proses di dalam steganografi, yaitu proses penyisipan pesan rahasia dan proses ekstraksi untuk mendapatkan kembali pesan rahasia tersebut.

B. Video [2]

Audio Video Interleave (AVI) adalah format file penyimpanan data- data multimedia. AVI diperkenalkan pertama kali oleh Microsoft pada bulan November 1992 sebagai bagian dari teknologi video dalam platform Microsoft Windows. Format file AVI dapat menyimpan data video dan audio dalam satu file yang memungkinkan memainkan kedua jenis data secara bersamaan. Dalam Tugas Akhir ini memakai avi jenis AVI uncompressed atau disebut juga AVI full frames. Suatu file multimedia dengan format AVI uncompressed memiliki informasi *frame* gambar yang disimpan dengan menggunakan format Bitmap tiga layer warna 8 bit, jadi untuk satu pixel data bitmap akan disimpan dalam wadah berukuran 24 bit.

C. Enhanced Least Significant Bit (ELSB)

Enhanced Least Significant Bit adalah modifikasi dari metode *Least Significant Bit*. Metode ini dapat dilakukan dengan dua cara. Pertama adalah dengan mengacak jumlah bit-bit dari *file host* yang digunakan untuk *embedding* pesan rahasia. Sementara cara yang kedua adalah dengan mengacak sampel *host* yang mengandung bit pesan rahasia selanjutnya. Pada ELSB, bit pada *host* yang digunakan untuk *embedding* pesan rahasia tidak selalu sama. Ashima Wadhawa [4] mengidentifikasi bahwa ELSB bekerja dengan baik melawan serangan dari *steganalysis* dan memiliki banyak keunggulan dibandingkan dengan LSB (*Least Significant Bit*). Proses ELSB membutuhkan waktu yang lebih lama karena memiliki ketentuan dalam pemilihan sampelnya.

Pemilihan lokasi bit untuk menyisipkan bit pesan dan pemilihan sampel yang digunakan untuk mentuk menyisipkan bit pesan memiliki aturan sebagai berikut.

Tabel 1 (a) Skema Pemilihan Letak Bit Pesan [3] (b) Skema Pemilihan Sampel [3]

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	Sampel yang berisi bit pesan berikutnya
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)

D. Mel-Frequency Warping (MFCC)

Mel-Frequency Cepstral Coefficient (MFCC) merupakan salah satu metode yang digunakan untuk melakukan ekstraksi ciri pada sebuah sinyal. Metode ini mengekstraksi ciri sinyal suara dengan mengadopsi cara kerja dari pendengaran manusia. [5]

Berikut adalah tahapan proses dalam *Mel-Frequency Cepstral Coefficient* [2]:

1. Pre-emphasis

Pre-emphasis merupakan *filter* yang digunakan untuk mempertahankan frekuensi-frekuensi tinggi dan mengurangi noise sehingga dapat meningkatkan kualitas sinyal.

2. Frame Blocking

Pada tahap ini, dilakukan proses *frame blocking* yaitu membaca sinyal suara per blok/*frame* yang terdiri dari S sampel karena kondisi sinyal audio yang berubah-ubah secara konstan.

3. *Windowing*

Proses *windowing* dilakukan pada setiap *frame* dengan tujuan untuk meminimumkan hilangnya informasi pada sinyal suara sehingga didapatkan suatu sinyal secara keseluruhan tanpa adanya informasi yang hilang.

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.

5. *Mel-Frequency Warping*

Tahap ini umumnya dilakukan dengan menggunakan *filterbank*. *Filterbank* adalah salah satu bentuk filter yang dilakukan dengan tujuan untuk mengetahui ukuran dari band frekuensi tertentu dalam sinyal suara. Pada MFCC digunakan *Mel Filterbank* yang menggunakan skala *mel*.

6. *Discrete Cosine Transform (DCT)*

Proses DCT ini menghasilkan koefisien *cepstral* dari sinyal suara.

7. *Cepstral Lifting*

Cepstral lifting menghaluskan spektrum hasil dari DCT sehingga dapat digunakan lebih baik untuk *pattern matching*.

E. Parameter Pengujian [2]

1. *Mean Square Error (MSE)*

Mean Square Error adalah parameter yang digunakan untuk menganalisis performansi sistem dengan melihat hasil kualitas *stego-video*. Dalam metode MSE ini dilakukan dengan cara mencari rata-rata nilai *error* antara citra *cover* dengan citra *stego*. Semakin besar nilai MSE yang didapat maka kualitas *stego-video* semakin buruk.

2. *Peak Signal to Noise Ratio (PSNR)*

PSNR merupakan tinjauan kualitas video secara objektif. PSNR adalah nilai tertinggi dari perbandingan daya sinyal dengan noise. Kualitas *stego-image* dapat dikatakan baik jika nilai PSNR-nya besar.

3. *Bit Error Rate (BER)*

BER (*Bit Error Rate*) merupakan parameter pengujian dimana bagus tidaknya sistem steganografi dan ekstraksi yang telah dibuat didasarkan pada benar atau tidaknya sistem dalam mengekstraksi bit-bit pesan yang telah dikirimkan. Parameter BER ini sangat menentukan bagus tidaknya sistem steganografi yang telah dibuat karena mengingat dari tujuan steganografi itu sendiri adalah menyampaikan pesan.

4. *Character Error Rate (CER)*

Character Error Rate (CER) adalah perbandingan jumlah karakter yang *error* dengan total karakter. CER merupakan parameter pengujian yang digunakan untuk melihat kualitas pesan yang disisipkan.

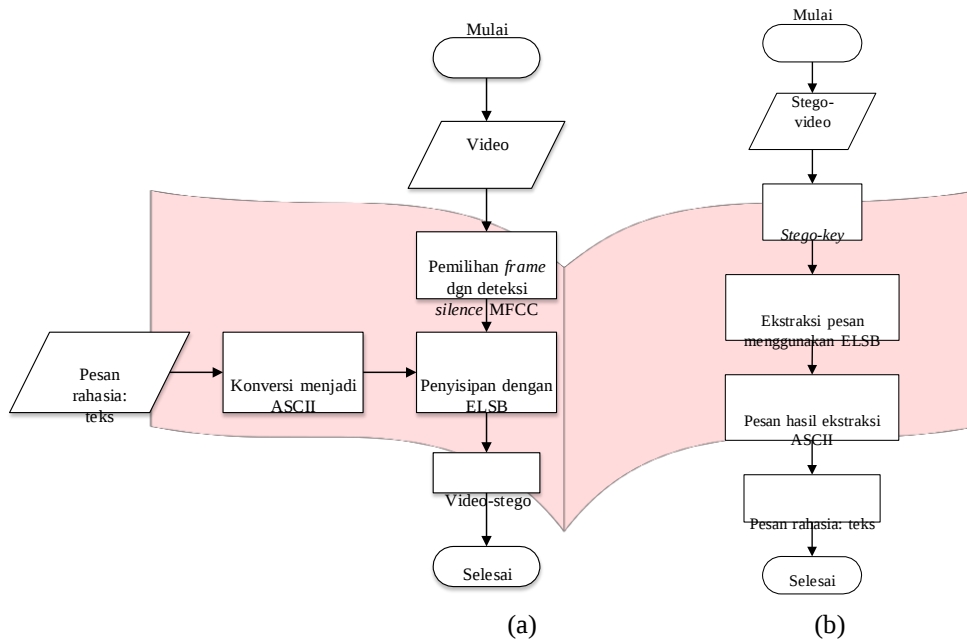
5. Waktu Komputasi

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

6. *Mean Opinion Score (MOS)*

Pengujian ini dilakukan dengan membandingkan kualitas 2 video, video asli dengan yang sudah disisipi pesan. Nilai yang digunakan pada MOS adalah dari 1 yang paling rendah dan 5 yang paling tinggi.

3. Perancangan Sistem



Gambar 1 Diagram Alir Proses (a) Penyisipan dan (b) Ekstraksi

Berdasarkan gambar 1 sistem yang dirancang pada tugas akhir ini adalah sistem steganografi dengan video sebagai *cover*. Penyisipan dilakukan di sisi pengirim dengan menyisipkan pesan rahasia berupa file teks dengan format .txt ke dalam video berformat .avi dengan metode *Enhanced Least Significant Bit*. Penyisipan dilakukan pada frame yang terdeteksi *silence Mel-Frequency Cepstral Coefficient*. Keluaran dari proses penyisipan ini berupa video-stego dimana terdapat pesan rahasia yang telah disisipkan. Kemudian video-stego dikirim ke penerima. Di sisi penerima dilakukan proses ekstraksi dengan bantuan stego-key untuk mengembalikan pesan rahasia berupa file teks dengan format .txt.

4. Pembahasan

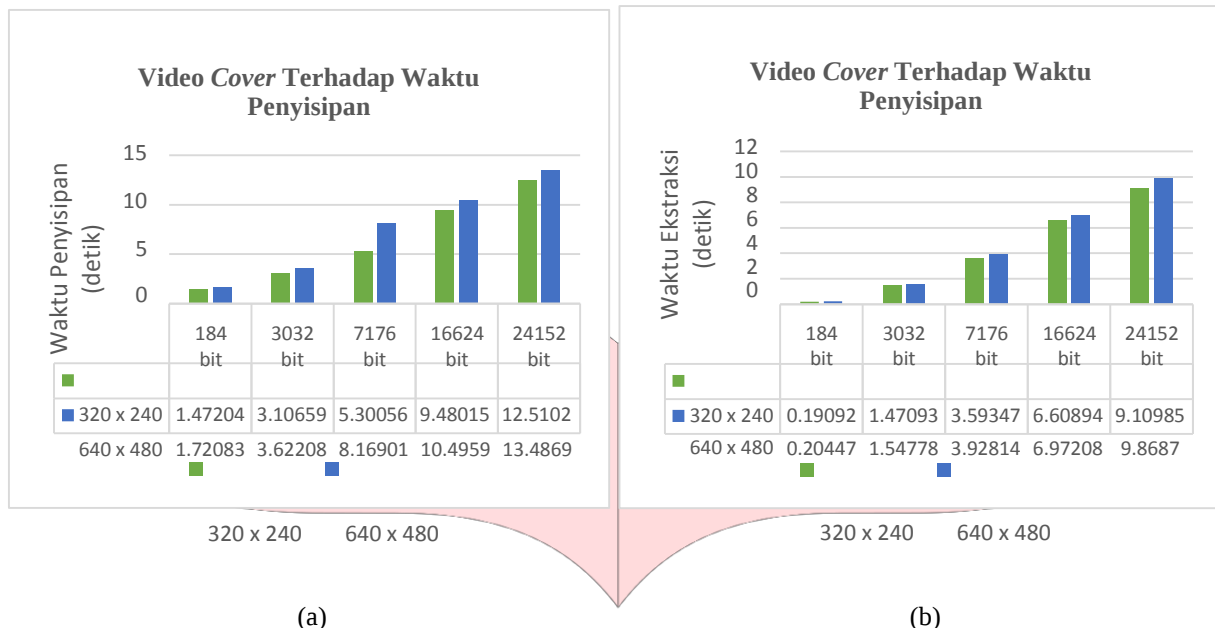
Pengujian sistem steganografi yang dirancang pada tugas akhir ini dilakukan dengan menyisipkan pesan rahasia ke dalam video *cover*. Pesan rahasia yang digunakan berupa teks dengan ukuran panjang 184 bit, 3032 bit, 7176 bit, 16624 bit dan 24152 bit. *Cover* yang digunakan adalah video dengan format .avi sebagai berikut:

Tabel 2 Video Cover

Kartun Dialog (doraemon.avi)	Kartun Lagu (cinderella_1.avi)	Kartun Lagu (cinderella_2.avi)	Musik (musik.avi)
Length: 00:00:07	Length: 00:00:07	Length: 00:00:07	Length: 00:00:07
Frame Width: 320	Frame Width: 320	Frame Width: 640	Frame Width: 640
Frame Height: 240	Frame Height: 240	Frame Height: 480	Frame Height: 480

A. Pengaruh Panjang Pesan dan Ukuran Video Cover Terhadap Waktu Komputasi

Pengukuran waktu komputasi akan dilakukan dengan menggunakan video *cover* ukuran 480 x 240 dan 640 x 480 dengan panjang pesan yang disisipkan adalah 184 bit, 3032 bit, 7176 bit, 16624 bit dan 24152 bit.



Gambar 2 (a) Pengaruh Panjang Pesan dan Ukuran Video Cover Terhadap Waktu Penyisipan (b) Pengaruh Panjang Pesan dan Ukuran Video Cover Terhadap Waktu Ekstraksi

Berdasarkan gambar tabel diatas, penyisipan 24152 bit pada cover video membutuhkan waktu komputasi hingga 12,5102 detik untuk proses penyisipan dan 9,10985 detik untuk ekstraksi. Sementara untuk panjang pesan 184 bit waktu penyisipan yang dibutuhkan hanya 1,47204 detik dan 0,19092 untuk ekstraksi. Dari hasil pengujian tersebut dapat diketahui bahwa semakin besar data yang disisipkan, maka semakin lama waktu komputasinya. Hal ini dikarenakan semakin panjang pesan rahasia maka semakin banyak pesan yang akan disisipkan, sehingga sistem membutuhkan waktu yang lebih lama untuk melakukan proses tersebut. Selain itu, semakin besar ukuran video cover maka semakin lama juga waktu komputasi yang diperlukan. Pesan dengan panjang 7176 bit disisipkan pada video dengan ukuran 320 x 240 membutuhkan waktu 5,30056 detik untuk penyisipan dan 3,59347 detik untuk ekstraksi. Sementara itu video dengan ukuran 640 x 480 membutuhkan waktu yang lebih lama yaitu 8,16901 detik untuk penyisipan dan 3,92814 detik untuk ekstraksi.

B. Pengaruh Jenis Video Cover Terhadap Daya Tampung Penyisipan

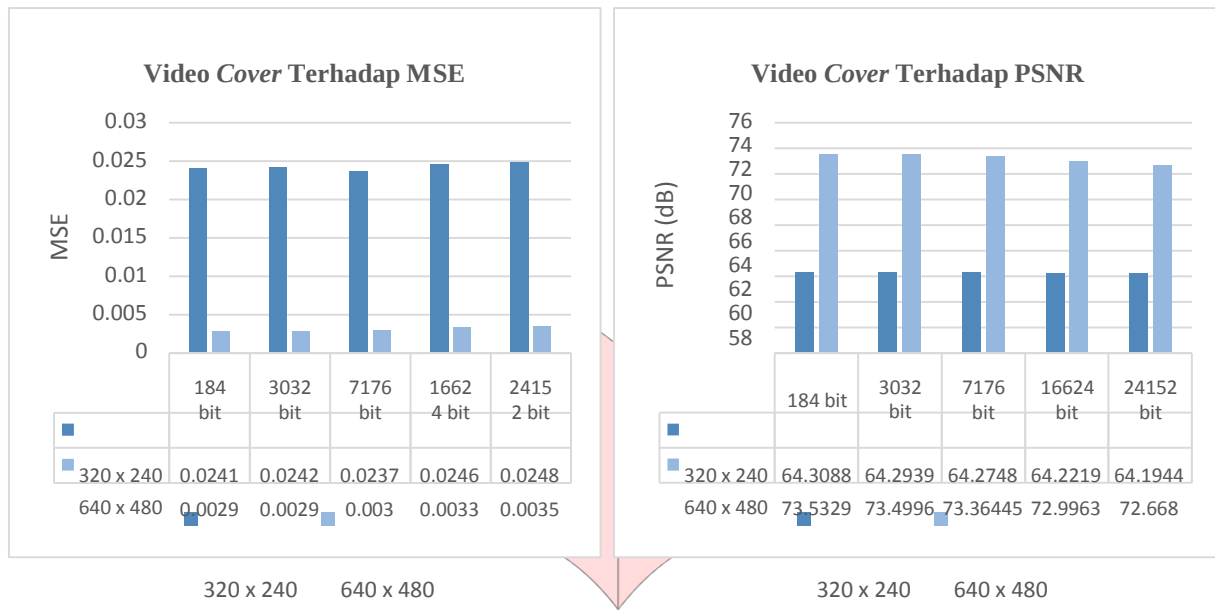
Pada pengujian ini dilakukan penyisipan pada 3 (tiga) jenis video cover yang berbeda, kartun dialog, kartun bernyanyi dan musik dengan panjang yang sama yaitu 7 detik dan jumlah *frame* total adalah 211.

	Kartun Dialog (Doraemon)	Kartun Lagu (Cinderella)	Musik
Jumlah Frame	26	23	61

Berdasarkan tabel diatas, dapat diketahui bahwa jenis video mempengaruhi jumlah daya tampung penyisipan pesan karena masing-masing video cover memiliki daya tampung yang berbeda tergantung dari seberapa lama terdapat sinyal *silence* yang terdeteksi dengan menggunakan MFCC.

C. Pengaruh Panjang Pesan dan Ukuran Video Cover Terhadap MSE dan PSNR

Pada pengujian ini, video kartun dialog akan diproses dengan sistem dengan ukuran video yang berbeda, Kartun Dialog (doraemon.avi) ukuran 320 x 240 dan Musik ukuran 640 x 480, dan juga dengan menyisipkan pesan rahasia sepanjang 184 bit, 3032 bit, 7176 bit, 16624 bit dan 24152 bit pada masing-masing video cover.



(a)

(b)

Gambar 3 (a) Pengaruh Panjang Pesan dan Ukuran Video Cover Terhadap MSE (b) Pengaruh Panjang Pesan dan Ukuran Video Cover Terhadap PSNR

Berdasarkan Gambar 3 (a) diatas, pesan yang disisipkan dan ukuran video dapat mempengaruhi nilai MSE. Semakin panjang pesan yang disisipkan maka semakin besar nilai MSE dan semakin besar ukuran video cover semakin besar pula nilai MSE yang didapat. Hal itu menunjukkan bahwa tingkat kemiripan video cover video stego semakin kecil dan tingkat kesalahan yang terjadi pada video *stego* meningkat. Sementara itu Gambar 3 (b) menunjukkan bahwa semakin panjang pesan yang disisipkan maka semakin kecil nilai PSNR dan semakin besar ukuran video, semakin besar pula nilai PSNR. Hal ini dikarenakan video 'musik.avi' memiliki ukuran video dan kapasitas penyisipan yang lebih besar dibandingkan 'doraemon.avi'. Nilai MSE berbanding balik dengan nilai PSNR. Semakin besar nilai MSE, maka semakin kecil nilai PSNR yang diperoleh, begitu pula sebaliknya. Semakin kecil nilai MSE maka semakin besar nilai PSNR yang diperoleh.

D. Pengaruh Panjang Pesan dan Ukuran Video Cover Terhadap Akurasi Pesan Terekstraksi

Pada pengujian ini dilakukan pada semua jenis video cover dengan menyisipkan pesan rahasia sepanjang 184 bit, 3032 bit, 7176 bit, 16624 bit dan 24152 bit pada masing-masing video.

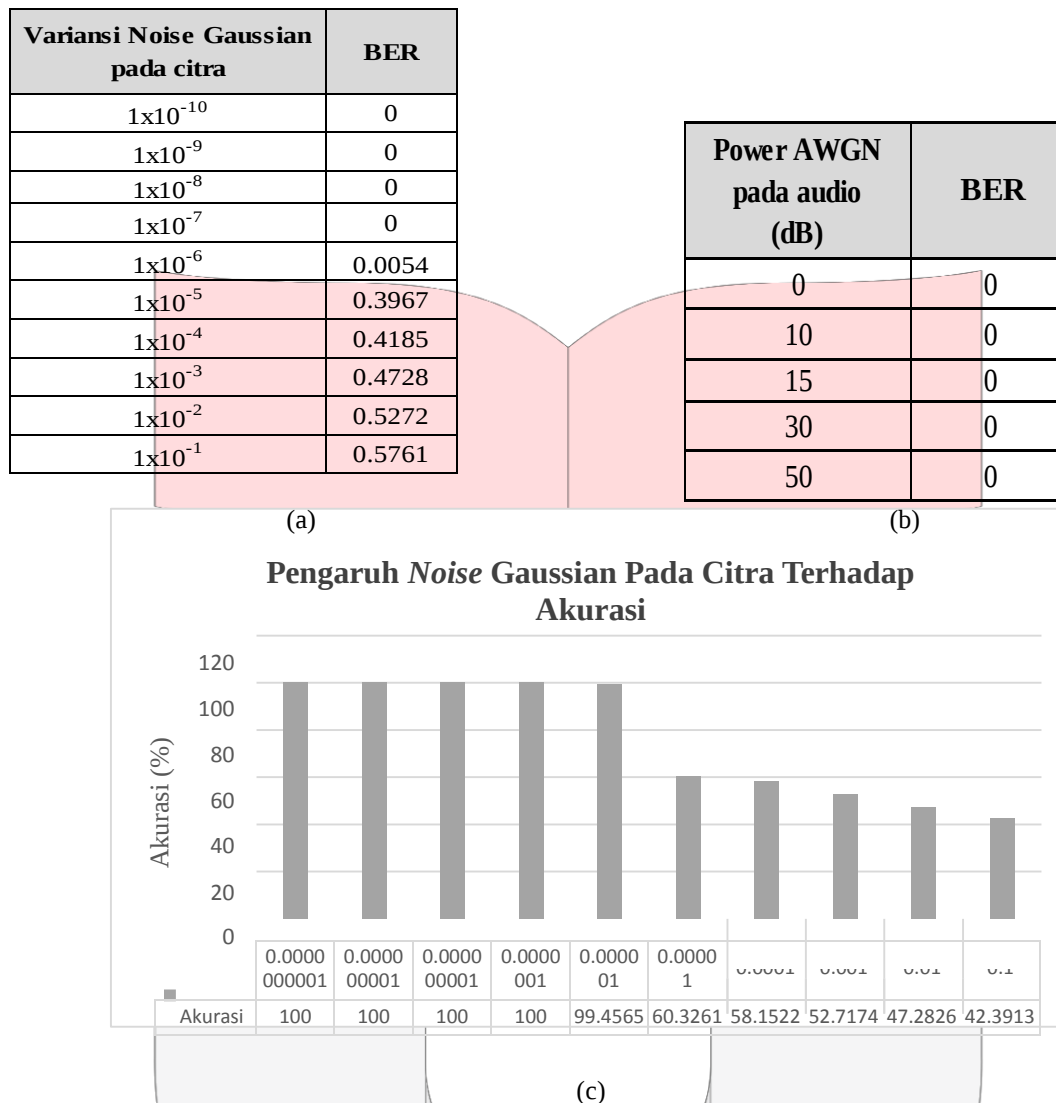
Nilai akurasi pesan terekstraksi pada sistem tetap 100% walaupun ukuran video cover dan panjang pesan berbeda-beda. Semakin besar ukuran video cover semakin banyak pula pesan yang dapat disisipkan, tetapi hal tersebut tidak mempengaruhi tingkat akurasi pesan terekstraksi.

E. Hasil Pengujian Terhadap Nilai BER dan CER

Setelah melakukan beberapa pengujian tanpa serangan *noise*, didapatkan nilai BER=0 dan CER=0. Hal ini menunjukkan bahwa tidak ada perbedaan nilai antara pesan asli sebelum disisipkan dengan pesan yang telah didapat saat proses ekstraksi.

F. Pengaruh Serangan Noise Gaussian Terhadap Nilai BER dan Akurasi

Pada pengujian ini ditambahkan *noise* Gaussian pada citra yang dilakukan pada saat mean=0 dengan variansi 1×10^{-10} sampai 1×10^{-1} pada video cover ukuran 320 x 240 yang telah disipi pesan sepanjang 184 bit. Selanjutnya akan dilakukan pengujian AWGN pada yang dilakukan pada saat mean=0 dengan power 0 dB, 10 dB, 15 dB, 30 dB dan 50 dB pada video cover ukuran 320 x 240 yang telah disipi pesan sepanjang 184 bit.



Gambar 4. (a) Pengaruh Noise Gaussian pada Citra Terhadap BER (b) Pengaruh Noise Gaussian pada Audio Terhadap BER (c) Pengaruh Noise Gaussian pada Citra Terhadap Akurasi

Pengujian dilakukan dengan menyisipkan pesan dengan panjang 184 bit pada video *cover* ukuran 320 x 240. Pada Gambar 4 (a) terlihat bahwa sistem dengan serangan *noise* Gaussian tahan hingga variansi 1×10^{-7} dengan nilai BER yang tidak sama dengan nol. Semakin besar nilai variansi *noise* Gaussian yang diberikan, maka kualitas video akan semakin buruk karena semakin banyak piksel yang berubah. Hal ini disebabkan karena jangkauan *noise* Gaussian yang semakin lebar. Hal tersebut juga ditunjukkan pada Gambar 4 (c) yaitu akurasi 100% hanya sampai pada variansi 1×10^{-7} . Dari gambar 4 (b) dapat terlihat bahwa *power* AWGN tidak mempengaruhi BER. Hal ini terjadi karena pada proses ekstraksi terdapat *stego-key* yang berisi letak pesan sehingga *noise* tidak mempengaruhi pembacaan nilai bit berdasarkan metode ELBS.

G. Uji Parameter MOS

Pengujian parameter MOS yang dilakukan bertujuan untuk melihat kualitas vide-stego jika diberi sisipan pesan dengan panjang yang berbeda. Panjang pesan yang disisipkan yaitu sepanjang 184 bit, 3032 bit, 7176 bit, 16624 bit dan 24152 bit. Video *cover* yang digunakan memiliki ukuran 320 x 240 dan 640 x 480.

Setelah melakukan survey terhadap 30 koresponden, didapatkan nilai rata-rata MOS sebagai berikut.

Tabel 3 Nilai Rata-rata MOS Terhadap Video Cover Ukuran 320 x 240

Ukuran Video	320 x 240				
Panjang Pesan (bit)	184	3032	7176	16624	24152
Nilai MOS Rata-rata	4.85714	4.78571	4.5	4.28571	4.21429

Tabel 4 Nilai Rata-rata MOS Terhadap Video Cover Ukuran 640 x 320

Ukuran Video	640 x 480				
Panjang Pesan (bit)	184	3032	7176	16624	24152
Nilai MOS Rata-rata	4.92857	4.85714	4.42857	4.32143	4.25

Berdasarkan Tabel 3 dan Tabel 4 diatas, diperoleh nilai rata-rata MOS video keseluruhan sebesar 4,5. Dari survey yang dilakukan dapat dilihat bahwa sistem steganografi yang dirancang memiliki kualitas baik. Dari hasil survey MOS, didapatkan nilai rata-rata MOS yang sama untuk video pertama dan kedua yaitu 4,5. Berdasarkan nilai yang didapatkan dapat disimpulkan bahwa kualitas video-stego adalah baik.

5. Kesimpulan

Dari hasil analisis pada pengujian didapatkan bahwa sistem yang dibuat mampu melakukan proses steganografi video menggunakan ELBS dengan pemilihan frame berdasarkan deteksi *silence* MFCC dengan baik. Berdasarkan tabel diatas, Panjang pesan dan ukuran video cover mempengaruhi waktu komputasi pada saat proses penyisipan dan ekstraksi pesan. Semakin panjang pesan yang disisipkan dan semakin besar ukuran video cover yang digunakan maka semakin lama waktu komputasi yang dibutuhkan untuk melakukan proses penyisipan dan ekstraksi pesan. Dari hasil pengujian stego video memiliki kualitas yang baik dan didapatkan PSNR mencapai 73,5329 dB. Nilai BER dan CER yang didapat dari seluruh pengujian dengan tanpa adanya serangan *noise* adalah 0. Hal ini menunjukkan bahwa tidak terdapat perbedaan antara pesan asli dan pesan yang didapat saat proses ekstraksi. Hasil parameter MOS yang didapatkan memiliki nilai rata-rata total sebesar 4,5 yang berarti kualitas video tersisipi dengan baik. Sistem yang dibuat tahan terhadap *noise* Gaussian dengan nilai mean=0 hingga variansi 1×10^{-7} .

Daftar Pustaka:

- [1] Berg G, Davidson, Ming-Yuan Dual, Paul G. 2003. *Searching For Hidden Message: Automatic Detection of Steganography*. Washington: Computer Science Departement, University at Albany.
- [2] Oktaviany, Arina Rizky. Dkk. 2008. "Implementasi dan Analisis Steganografi Video Berbasis Wavelet". Jurusan Teknik Elektro, Institut Teknologi Telkom, Bandung.
- [3] Muda, L, Mumtaj, B., & Elamvazuthi, L, 2010, *Voice Recognition Algorithm using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DWT) Techniques*, Journal of Computing Volume 2, Issue 3, March 2010.
- [4] Wadhwa, Ashima. 2014. *A Survey on Audio Steganography Techniques for Digital Data Security*. International Journal of Advanced Research in Computer Science and Software Engineering.
- [5] 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