DETEKSI PESAN DENGAN METODE DIFFERENCE RATIO STEGANALYSIS DAN KLASIFIKASI K-NN UNTUK STEGANOGRAFI YANG TERSISIPI PESAN SECARA MDCT

MESSAGE DETECTION BY DIFFERENCE RATIO STAGANALYSIS AND K-NN METHOD FOR STEGANOGRAPHY AUDIO THAT INCLUDES MESSAGE BY MDCT

Jovita Michaela Nariyari[1], Iwan Iwut Tritoasmoro[2], Nur Ibrahim[3]

[1,2,3] Universitas Telkom, Bandung
jovitamichaela@student.telkomuniversity.ac.id[1],
iwaniwuttritoasmoro@telkomuniversity.ac.id[2], **nuribrahim@telkomuniversity.ac.id**[3]

Abstrak

Pertukaran informasi yang mudah pada era digital ini, turut mempengaruhi ilmu steganografi. Ilmu ini sewaktu-waktu dapat menjadi tindak kejahatan yang disalahgunakan oleh oknum tidak bertanggung jawab. Salah satu kasusnya adalah tindak terorisme yang masih marak terjadi. Berlatarkan masalah tersebut, maka diperlukan anti-steganografi yang mampu mendeteksi informasi yang dicurigai, ilmu anti-steganografi tersebut dikenal dengan istilah steganalisis.

Difference Ratio Steganalysis* dan K-Nearest Neighbour merupakan metode yang digunakan dalam merancang sistem steganalisis. Pada penelitian ini, sistem yang dirancang bertujuan untuk mendeteksi keberadaan dan posisi pesan pada sebuah file audio yang tersteganografi oleh *Modified Discrete Cosine Transform.

Berdasarkan pengujian yang dilakukan, didapatkan hasil akurasi terbaik 77,5% dalam mendeteksi keberadaan pesan dan 100% dalam mendeteksi posisi pesan. Hasil tersebut didapatkan dengan mengunakan beberapa parameter yaitu ukuran frame, nilai K, jarak K-NN, dan ukuran pesan.

Kata Kunci: Steganografi, Steganalisis, DR Steganalysis, MDCT, K-NN

Abstract

The easy exchange of information in this digital era, also has affected the steganography. It any time can become a crime that is misused by irresponsible guy. One of that is terrorism, which is that still happening. Based on the problem, anti-steganography is required to detecting suspected information, anti-steganography called steganalysis.

Difference Ratio Steganalysis and K-Nearest Neighbour are methods used to design the steganalysis system. In this research, the system purposes to detect the existence and position of message in a Modified Discrete Cosine Transform steganographed audio.

Based on the test, got the best result 77,5% to detect message existence and 100% to detect message position. These result got while using parameters i.e. frame size, K value, K-NN distance, and size of message.

Keyword: Steganography, Steganalysis, DR Steganalysis, MDCT, K-NN

1. Pendahuluan

Ilmu steganografi pertama kali digunakan oleh rakyat Yunani untuk melawan Kerajaan Persia. Rakyat Yunani menggunakan kepala budak menjadi media penyisipan untuk menyampaikan pesan [1]. Seiring perkembangan teknologi, tentunya cara ini tidak dipraktikan lagi. Perkembangan teknologi yang semakin pesat, telah menghantarkan kita pada era digital, dimana semua informasi dapat dengan mudah diakses oleh siapapun. Media informasi pun beragam mulai dari teks, citra, audio, bahkan video. Tidak dapat dipungkiri penyalahgunaan ilmu steganografi dan kemajuan teknologi bisa terjadi. Salah satu contohnya adalah tindakan terorisme yang masih marak terjadi, dengan ilmu ini pelaku dapat menyisipkan titik koordinat lokasi, peta, dan foto target sebagai pesan

rahasia. Berdasarkan hal tersebut, diperlukan sebuah sistem yang mampu mendeteksi media-media yang dicurigai mengandung informasi-informasi tersebut. Steganalisis hadir sebagai tindakan balik terhadap steganografi yang dapat diartikan sebagai seni komunikasi tersembunyi. Steganalisis memiliki tiga tingkatan yakni: mendeteksi, mengekstraksi, dan menonaktifkan atau menghancurkan pesan yang tersembunyi [2].

Pada beberapa tahun sebelumnya telah dilakukan penelitian yang berhubungan dengan steganalisis. Pada penelitian tersebut [3], dilakukan pendeteksian eksistensi pesan terhadap audio .wav yang tersisipi pesan secara DCT, DWT, dan ELSB. Pada penelitian ini digunakan metode *Discrete Cosine Transform* dan *Principal Component Analysis* sebagai metode pendeteksian dalam simulasi. Dalam tugas akhir ini digunakan metode *Diffrence Ratio Steganalysis* dan klasifikasi *K-Nearest Neighbour* untuk mendeteksi eksistensi dan posisi pesan yang tersisipi secara *Modified Discrete Cosine Transform*. Penggunaan metode *DR Steganalysis* dalam penelitian ini dimaksud untuk mengubah karakteristik dari nilai statistik yakni perbedaan nilai mean positif dan negatif pada audio sehingga eksistensi pesan dapat terdeteksi. Klasifikasi *K-Nearest Neighbour* digunakan untuk mengkalsifikasikan audio yang memiliki sisipan sehingga dapat dideteksi kembali untuk mencari posisi pesan pada audio. Metode ini dapat mengklasifikasikan secara tepat karena memiliki garis keputusan kelas nonlinear.

2. Dasar Teori

2.1 Steganografi

Steganografi merupakan sebuah seni dan ilmu dalam berkomunikasi secara tersembunyi, dengan kata lain sebuah teknik penyembunyian informasi didalam sebuah media yang tidak dapat diketahui oleh siapapun, kecuali pengirim dan penerima. Steganografi terdiri atas 2 suku kata yang berasal dari bahasa Yunani yakni steganos (wadah) dan graphein (tulisan), bila digabungkan memiliki arti tulisan yang tertutup [4]. Cara kerja steganografi adalah dengan menyisipkan pesan dengan algoritma tertentu pada sebuah media lalu dikirimkan ke sisi penerima [5]. Perkembangan steganografi hingga saat ini memiliki 5 tipe cover sebagai media untuk menyembunyikan pesan yaitu steganografi citra, steganografi teks, steganografi audio, steganografi video, dan steganografi *protocol* [6].

2.2 Steganalisis

Steganalisis adalah ilmu untuk mendeteksi keberadaan pesan yang telah disembunyikan menggunakan metode steganografi tertentu. Menurut [7] [8], steganalisis secara luas memiliki dua kategori yaitu: algorithm spesific atau targeted dan universal method. Pada algorithm spesific, penyerang dianggap mengetahui metode steganografi yang digunakan sedangkan, pada universal method merupakan kebalikan dari kategori sebelumnya. Selain itu, terdapat dua model penyerang yakni penyerang aktif dan pasif. Penyerang pasif hanya akan mendengarkan audio tanpa peduli akan isi dari pesan tersembunyi tersebut, sebab tujuan penyerang adalah untuk mendeteksi keberadaan pesan saja. Sedangkan, penyerang aktif akan berusaha menjaga file tersebut karena ingin mengetahui isi dari pesan yang tersembunyi pada audio tersebut. Pada steganalysis juga terdapat istilah active steganalysis dan passive steganalysis. Berdasarkan metode penyerangan terdapat 6 kategori utama dalam steganalisis yaitu *visual steganalysis*, *signature steganalysis*, *statistical steganalysis*, *spread spectrum steganalysis*, *transform domain steganalysis*, dan *universal or blind steganalysis* [9].

2.3 Audio Digital

Audio digital adalah audio yang telah diubah asal sinyalnya dari analog menjadi digital dengan mengkodekan angka biner hasil pengubahan sinyal dengan menggunakan frekuensi sampling. Audio berhubungan dengan indra manusia yaitu indra pendengaran, yang mana sistem pendengaran manusia ini memiliki kepekaan yang baik terhadap berbagai jenis suara baik yang bersumber dari manusia sendiri maupun benda mati. Karakteristik audio salah satunya jarak frekuensi, untuk manusia sendiri berada pada frekuensi 20 Hz-20KHz sesuai dengan tekanan yang dirasakan oleh gendang telinga [10]. Salah satu format audio digital adalah WAV. WAV atau WAVE merupakan salah satu standar format audio digital yang sering digunakan saat ini. Standar format ini dikembangkan oleh Microsoft dan IBM pada tahun 1991. Format ini merupakan contoh dari Resource Interchange File Format (RIFF) dan berbasis pada ”chunk” [11].

2.4 *Difference Ratio Steganalysis*

Difference Ratio merupakan sebuah metode steganalisis yang digunakan untuk mendeteksi ada tidaknya objek pada sebuah pesan yang dicurigai dengan menggunakan perbedaan nilai rasio. Metode ini bekerja sesuai dengan prinsip steganalysis bahwa nilai mean fragmen positif (NPMF) dan negatif (NNMF) setelah dilakukan steganografi hampir sama, sehingga perbedaan rasio sebuah informasi digital sebelum dilakukan steganografi lebih kecil dibandingkan dengan perbedaan rasio setelah dilakukan proses steganografi. Respresentasi rumus dari metode DR menurut [12].

$$DR = \frac{|NPMF - NNMF|}{(min(NPMF, NNMF))} \quad (1)$$

Untuk mengetahui ada tidaknya informasi yang disisipkan, metode ini melakukan proses steganografi kembali terhadap file yang telah disteganografikan sebelumnya, secara singkat metode ini melakukan dua kali proses steganografi untuk melihat perbadaan rasio antara penyisipan pertama dan kedua.

$$DR = \frac{DR2}{DR1} \quad (2)$$

2.5 *K-Nearest Neighbour*

Algoritma K-Nearest Neighbor merupakan sebuah metode pengklasifikasian yang mancari jarak terdekat dari nilai yang akan dievaluasi (titik queri) dengan tetangga terdekatnya dalam suatu data [13][14]. [15] Dasar algoritma ini adalah nilai k sebagai ciri nilai tetangga terdekat dalam sebuah data, dimana sebagian data sampel masuk dalam kategori tertentu yang telah ditetapkan. Secara umum, algoritma ini digunakan untuk pengklasifikasian, dimana jika pada sebuah nilai k muncul data sampel yang banyak maka nilai tersebut yang digunakan sebagai prediksi. Untuk hasil yang signifikan digunakan perhitungan jarak. Berikut adalah beberapa perhitungan jarak yang sering digunakan, antaralain:

A. Cityblock Distance, fungsi ini merupakan penjumlahan dari nilai selisih absolut dari jarak antara titik p dan q.

$$D(p,q) = \sum_{i=1}^{n} |p_i - q_i| \quad (3)$$

B. Euclidean Distance, perhitungan jarak secara garis lurus antara dua titik. Jarak ini juga disebut phytagoras distance.

$$D(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (4)$$

C. Cosine Distance, jarak dari sudut antar titik di dan dj (diperlakukan seperti vector).

$$\cos(d_i, d_j) = \frac{\sum_{k=1}^{n} a_{ik} a_{ik}}{\sqrt{\sum_{k=1}^{n} a^2{}_{ik}} \sqrt{\sum_{k=1}^{n} a^2{}_{jk}}} \quad (5)$$

D. Correlation Distance, perurutan nilai yakni 1-b, dimana b merupakan korelasi antara titik-titik sampel.

$$S_{ij} = \frac{\sum_{k=1}^{n} (X_{ik} - \overline{X_I})(X_{jk} - \overline{X_J})}{\left[\sum_{k=1}^{n} (X_{ik} - \overline{X_I})^2 \sum_{k=1}^{n} (X_{jk} - \overline{X_J})^2\right]} \quad (6)$$

2.6 *Modified Discrete Cosine Transform*

Modified Discrete Cosine Transform (MDCT) adalah salah satu teknik transformasi sinyal yang mengubah sinyal dari domain waktu menjadi domain frekuensi. MDCT merupakan cabang dari

Discrete Cosine Transform yang berbasis pada persamaan DCT-IV. Teknik transformasi ini, memiliki basis fungsi transformasi yang akan membuat overlap pada batas antarblok. Dalam hal ini jika sebuah frame memiliki panjang N, makan teknik ini akan memproses frame sebanyak 2N untuk menghasilkan beberapa N koefisien [16].

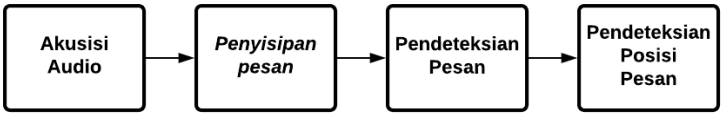
2.7 *Quantization Index Modulation*

Quantization Index Modulation (QIM) adalah salah satu teknik untuk menyembunyikan data dalam sebuah media. Cara kerja teknik ini dengan membagi sinyal menjadi beberapa bagian [16].

3. Perancangan Sistem

3.1. Desain Sistem

Dalam penelitian tugas akhir ini, sistem dirancang menggunakan aplikasi MATLAB R2018a yang mengacu pada gambar 1.

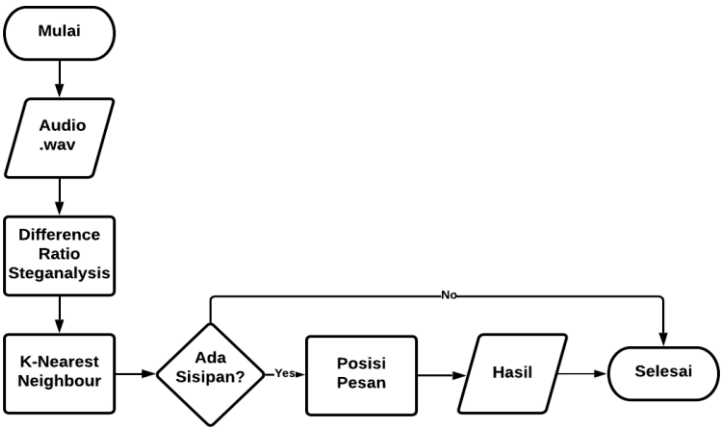


Gambar 1. Model Desain Umum Sistem

Pada tahap pertama, dikumpulkan audio .wav berisi percakapan sebanyak 20 audio yang dipisahkan menjadi 10 audio data latih dan 10 audio data uji. Tahap selanjutnya adalah melakukan penyisipan terhadap audio menggunakan metode MDCT. Pesan yang disisipkan berupa citra biner. Kemudian audio yang telah disisipi maupun tidak disisipi di deteksi untuk menganalisis kemampuan sistem yang telah dirancang, pada tahap ini digunakan metode DR Steganalysis. Kemudian audio yang dideteksi memiliki sisipan pesan, dideteksi kembali untuk mengetahui posisi pesan yang tersembunyi.

3.2 Desain Perancangan Sistem Steganalisis

Pada penilitian ini, proses alur sistem steganalisis dijabarkan pada gambar 2.



Gambar 2. Model Desain Sistem Steganalisis

Langkah pertama yang dilakukan adalah memilih audio yang akan diproses. Sesuai dengan cara kerja DR Steganalysis, audio yang telah dipilih disteganografikan menggunakan metode MDCT untuk mendapatkan nilai NPMF1, NNMF1, dan DR1. Audio yang telah disteganografikan kemudian *re-steganography* menggunakan metode yang sama untuk mendapatkan nilai NPMF2, NNMF2, dan DR2. Setelah itu maka akan didapatkan nilai Q yang diperoleh seperti rumus (2). Tahap selanjutnya adalah dilakukan pengklasifikasian audio menggunakan metode K-NN dimana jika ditemukan

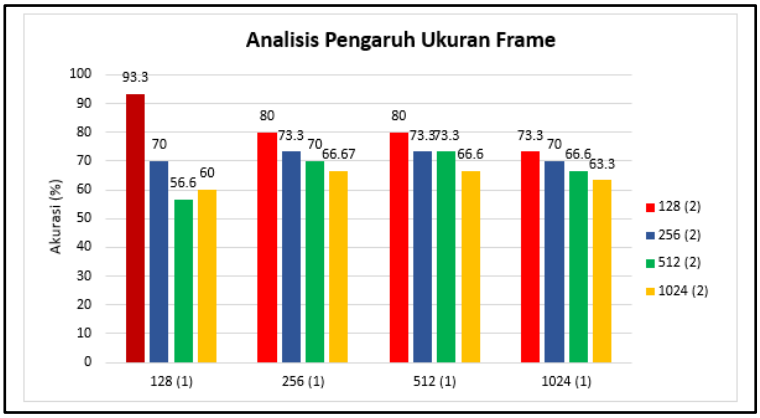
sisipan pesan maka audio tersebut diproses Kembali untuk mengetahui posisi sisipan pesan. Proses selesai apabila audio yang dideteksi memiliki sisipan pesan dan posisi sisipan atau tidak memiliki keduanya.

4. Hasil Analisis

A. Pengujian Terhadap Sistem Steganalisis

a.1. Analisis pengaruh ukuran frame

Pada pengujian ini dilakukan berdasarkan pengaruh ukuran frame. Variasi ukuran frame yang digunakan yaitu 128, 256, 512, dan 1024. Pengujian dilakukan dengan menyisipkan pesan menggunakan ukuran frame yang sama atau berbeda pada penyisipan pertama dan penyisipan yang kedua.

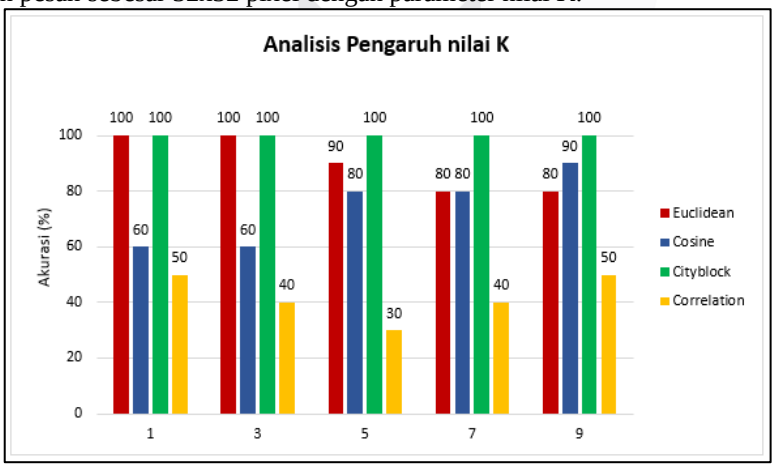


Gambar 3. Grafik Akurasi Pengujian Berdasarkan Pengaruh Ukuran Frame

Pada proses ini didapatkan tingkat akurasi tertinggi ketika menggunakan ukuran frame 128 *to* 128 yaitu sebesar 93,3%.

a.2. Analisis pengaruh nilai K

Audio yang menggunakan ukuran frame 128 *to* 128 kemudian diproses kembali menggunakan ukuran sisipan pesan sebesar 32x32 pixel dengan parameter nilai K.

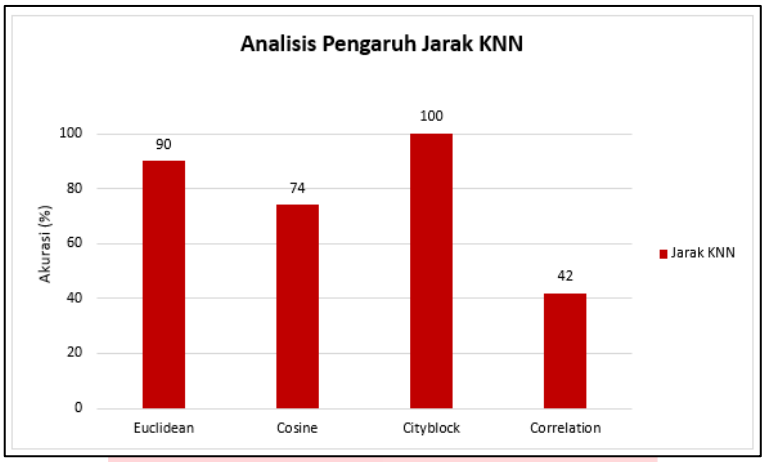


Gambar 4. Grafik Akurasi Pengujian Berdasarkan Pengaruh Nilai K

Berdasarkan gambar 4., didapatkan akurasi rata-rata tertinggi ketika menggunakan K=9 yaitu 80%.

a.3. Analisis pengaruh jarak K-NN

Pada pengujian ini digunakan jarak K-NN sebagai parameter yang bertujuan untuk menentukan pengaruh jarak yang sesuai pada sistem. Jenis jarak yang digunakan adalah *euclidean*, *cosine*, *cityblock*, dan *correlation*.

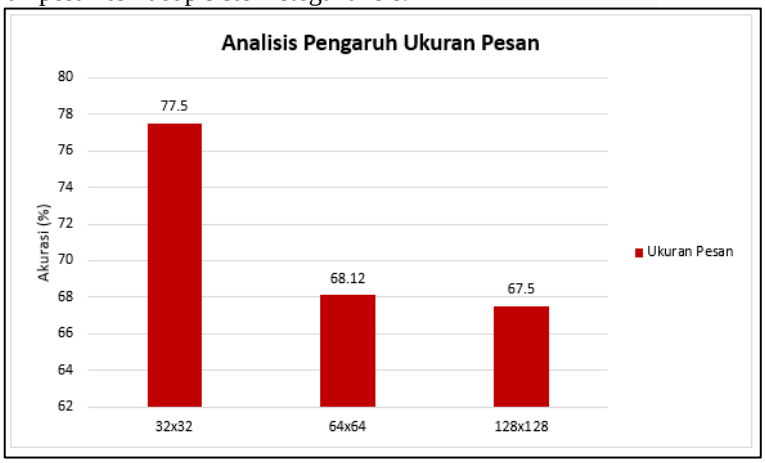


Gambar 4. Grafik Akurasi Pengujian Berdasarkan Pengaruh Jarak K-NN

Hasil yang diperoleh pada pengujian ini adalah penggunaan jenis jarak *cityblock* lebih sesuai digunakan pada sistem steganalisis ini.

a.4. Analisis pengaruh ukuran pesan

Pada pengujian ini digunakan ukuran pesan (citra) sebagai parameter. Variasi ukuran pesan yang digunakan adalah citra berukuran 32x32,64x64, dan 128x128. Audio yang digunakan sebagai *host* adalah audio dengan ukuran frame 128 *to* 128. Tujuan dari pengujian ini adalah untuk mengetahui pengaruh ukuran pesan terhadap sistem steganalisis.



Gambar 5. Grafik Akurasi Pengujian Berdasarkan Pengaruh Ukuran Pesan

Dari hasil pada gambar 5, dapat diketahui bahwa ukuran pesan turut mempengaruhi performansi sistem. Sistem memiliki akurasi yang baik ketika ukuran pesan sebesar 32x32.

B. Pengujian Terhadap Sistem Deteksi Posisi Pesan

Data yang digunakan pada pengujian deteksi posisi ini adalah data yang terindikasi memiliki sisipan, dimana data ini sebelumnya telah diuji dan mendapatkan akurasi tertinggi. Berikut merupakan 10 data audio tersebut, dimana 5 diantaranya terindikasi memiliki sisipan.

Tabel 1. Data Uji

Daftar Audio	Status Sisipan	Posisi Sisip (pada *n* ke - …)
audio241.wav	embed	6
audio242.wav	not embed	-
audio243.wav	embed	11
audio244.wav	not embed	-
audio245.wav	embed	6
audio246.wav	not embed	-
audio247.wav	embed	11
audio248.wav	not embed	-
audio249.wav	embed	6
audio250.wav	not embed	-

Pada pengujian ini diperoleh hasil akurasi sistem deteksi posisi sebesar 100%, dimana sistem dapat mendeteksi audio yang memiliki sisipan maupun tidak memiliki sisipan. Berikut adalah tabel hasil dari pengujian ini yang dipaparkan pada Tabel 2.

Tabel 2. Data Hasil Uji

Daftar Audio	Status Sisipan	Posisi Sisip (pada *n* ke - …)	Hasil
audio241.wav	embed	6	benar
audio242.wav	not embed	-	benar
audio243.wav	embed	11	benar
audio244.wav	not embed	-	benar
audio245.wav	embed	6	benar
audio246.wav	not embed	-	benar
audio247.wav	embed	11	benar
audio248.wav	not embed	-	benar
audio249.wav	embed	6	benar
audio250.wav	not embed	-	benar

5. Kesimpulan dan Saran

A. Kesimpulan

1. Akurasi yang diperoleh untuk acc-exist sebesar 77,5% dan untuk acc-pos sebesar 100%.
2. Kondisi terbaik didapatkan ketika menggunakan ukuran frame 128 *to* 128, nilai K=9, jarak *Cityblock*, dan ukuran pesan 32x32 pixel.
3. Parameter-parameter yang mempengaruhi sistem ini adalah ukuran frame, nilai K, jenis jarak K-NN, dan ukuran pesan yang disisipi.

B. Saran

1. Menggunakan metode steganalisis yang lebih bervariasi, agar dapat mengekstraksi isi pesan
2. Menggunakan sisipan yang lebih bervariasi, misalkan citra berwarna
3. Sistem yang dibuat dapat diimplementasikan secara realtime seperti android atau ios.

Referensi

[1] A. Siper, R. Farley and C. Lombardo, "The Rise of Steganography," *Proceedings of Student/Faculty Research Day, CSIS, Pace University*, 2005.

[2] T. Qian and S. Manoharan, "A Comparative Review of Steganalysis Techniques," *IEEE 2nd International Conference on Information Science and Security (ICISS) 2015*, 2016.

[3] A. E. Mahareni, "Simulasi Steganalisis Audio Digital Berbasis Discrete Cosine Transform dan Principal Component Analysis," *Tugas Akhir. Jurusan Teknik Telekomunikasi. Universitas Telkom: Bandung.*, 2014.

[4] A. M. Ferreira, "An Overview On Hiding and Detecting Stego-Data In Video Streams," *Research Project II. Dept. System & Networking Engineering. University of Amsterdam: Netherlands.*, 2015.

[5] Z. S. Madlool, S. A. Faris and A. M. Hussein, "A Review of Various Steganography Techniques in Cloud Computing," *University of Thi-Qar Journal of Science (UTSci)*, vol. 7, no. 1, pp. 1-5, 2019.

[6] P. Sharma and P. Kumar, "Review of Various Image Steganography and Steganalysis Techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 6, no. 7, pp. 152-159, 2016.

[7] H. Ghasemzadeh and M. H. Kayvanrad, "Comprehensive Review of Audio Steganalysis Method," *IET Signal Processing* , vol. 12, no. 6, pp. 673-687, 2018.

[8] R. Nouri and A. Mansouri, "Blind Image Steganalysis Based on Reciprocal Singular Value Curve," *Iranian Conference on Machine Vision and Image Processing, MVIP*, Vols. 2016-Febru, no. 1, pp. 124-127, 2016.

[9] K. Karampidis, E. Kavallieratou and G. Papadourakis, "A Review of Image Steganalysis Techniques for Digital Forensics," *Journal of Information Security and Applications*, vol. 40, no. May, pp. 217-235, 2018.

[10] M. Darji, *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 2, no. 3, 2017.

[11] S. Wilson and M. Bosi, "WAVE PCM Soundfile Format", dalam http://tiny.systems/software/soundProgrammer/WavFormatDocs.pdf, diakses pada 9 Maret 2020.

[12] M. Lei, Y. Yang, X. Niu and S. Luo, "Audio Steganalysis in DCT Domain," *Frontiers of Electrical and Electronic Engineering in China*, vol. 5, no. 2, pp. 203-206, 2010.

[13] J. Williams and Y. Li, "Comparative Study of Distance Functions for Nearest Neighbors," *Advanced Techniques in Computing Sciences and Software Engineering*, no. August 2018, 2010.

[14] A. F. Ryamizard, "Deteksi Nada Tunggal Alat Musik Kecapi Bugis Makassar Menggunakan Metode Mel Frequency Cepstral Coefficient (MFCC) dan Klasifikasi K-Nearest Neighbour (KNN)," *Tugas Akhir. Jurusan Teknik Telekomunikasi. Universitas Telkom: Bandung.*, 2018.

[15] W. Zhang, X. Chen, Y. Liu and Q. Xi, "A Distributed Storage and Computation K-Nearest Neighbor Algorithm Based Cloud-Edge Computing for Cyber-Physical-Social Systems," *IEEE Access*, vol. 8, pp. 50118-50130, 2020.

[16] D. Rosari, "Analisis Optimasi Steganografi Audio Berbasis MDCT dengan Algoritma Genetika," *Tugas Akhir. Jurusan Teknik Telekomunikasi. Universitas Telkom: Bandung.* , 2017.