CHAPTER 4

ENCODING PROCESS IN VIDEO STEGANOGRAPHY

The encoding process is the first and most crucial step in video steganography. It describes how to securely include private information in a cover video so that viewers cannot see it and it is impervious to manipulation or attack. This project's encoding process consists of three primary steps: audio steganography for metadata embedding, image steganography for secret text embedding, and deep learning-based frame selection. The outcome is a fully functional stego-video that preserves the audio and visual quality while hiding sensitive information.

4.1 Frame Selection Using DNN

Instead of embedding data into random frames, this project employs a Deep Neural Network (DNN) trained to assess visual suitability for steganography to intelligently choose five ideal frames from the video.

4.1.1 Steps in Frame Selection:

1. **Video-Audio Separation**:

   The input video is divided into:

   a. **Muted Video** (no audio)

   b. **Audio Track**

2. **Feature Analysis**:

   Each video frame is passed through a trained DNN that analyzes:

   a. **Blurriness**: Determines clarity of the frame.

   b. **Contrast**: Measures the color intensity difference.

   c. **Entropy**: Evaluates the randomness of pixel information.

   d. **Edge Information**: Indicates the presence of defined object boundaries.

   e. **Dominant Color**: Analyzes color patterns and uniformity.

   f. **JPEG Compression Artifacts**: Checks for distortions.

   g. **Keyframe Approximation**: Detects frames critical for visual flow

Frame Index	Blurriness	Contrast	Entropy	Edge Density	Color Ratio	DCT Energy	Frame Difference	Predicted Unsuitability Score
320.00	3193.2	58.50	7.63	50.57	0.54	922.55	11.69	4256.4
335.00	3191.4	58.54	7.63	50.37	0.54	920.06	11.69	4252.0
330.00	3194.0	58.47	7.63	50.34	0.54	916.16	11.70	4250.8
315.00	3176.0	58.57	7.63	50.37	0.54	927.00	11.64	4253.5
325.00	3186.0	58.49	7.63	50.46	0.54	916.82	11.68	4243.4

Table 4.1 Frame Selection metrics

3. **Scoring & Selection**:

   The DNN assigns an unsuitability score to each frame, and the five frames with the lowest scores are selected to act as steganographic carriers. Their indexes are saved in a metadata text file, which is essential for the decoding procedure.

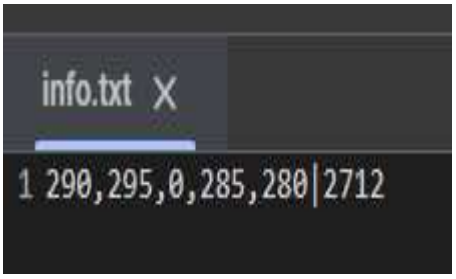

Fig 4.1.1 Metadata Text

Fig 4.1.2 Info Text

Why Use DNN for Frame Selection?

1. It enhances imperceptibility by choosing frames less sensitive to human eyes.
2. Reduces the risk of steganalysis by embedding data in visually insignificant regions.
3. Ensures a consistent, learned strategy instead of random or manual selection.

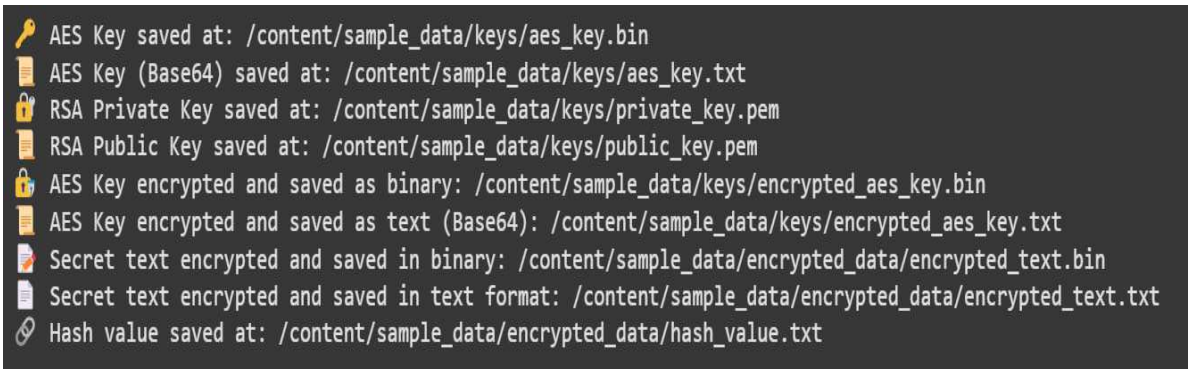

Fig 4.1.3 Use of DNN for Frame selection

4.2 Audio Steganography for Frame Index

Once the least appropriate frames have been selected and stored in the metadata file, the next step is to safely hide the metadata. For this, audio steganography is employed. The following are some of the actions involved:

4.2.1 Hybrid Hashing: Once the image indices have been established, the next step is to generate a unique and secure hash value that maintains the integrity and authenticity of the data. This is achieved by creating a hybrid hashing mechanism by fusing two strong and well-liked cryptographic hash functions, MD5 and SHA-256.

1. **MD5(MessageDigestAlgorithm5):**
   The image indices are first subjected to the MD5 algorithm. The fast hashing algorithm MD5 produces a hash value that is 128 bits (16 bytes). When used alone, MD5 is vulnerable to collision attacks even though it is useful in creating a condensed representation of the data. Because of this, modern systems do not consider MD5 alone to be secure.
2. **SHA-256(SecureHashAlgorithm256-bit):**
   To overcome the limitations of MD5, the output of the MD5 hash is further processed using the SHA-256 algorithm, which generates a 256-bit (32-byte) hash value. SHA-256 is a member of the SHA-2 family, which is well known for its high security and resistance to collisions. By successively combining the security strength of SHA-256 and the speed of MD5, this hybrid hashing technique exploits both. Because the final hash value generated ensures that even the smallest alteration to the indices would result in a completely different hash, data integrity is preserved.

Fig 4.2.1 Image of Hash value generated

4.2.2 AES Encryption:

1. The plain text secret message is first encrypted using AES, a symmetric key encryption algorithm that is well-known for its efficiency and security. AES uses fixed block sizes of 128 bits and can handle key lengths of 128, 192, or 256 bits.
2. The message is encrypted using a random AES symmetric key, which turns it into ciphertext that cannot be decrypted without the decryption key. The message will be safely converted into an unintelligible format using AES encryption.

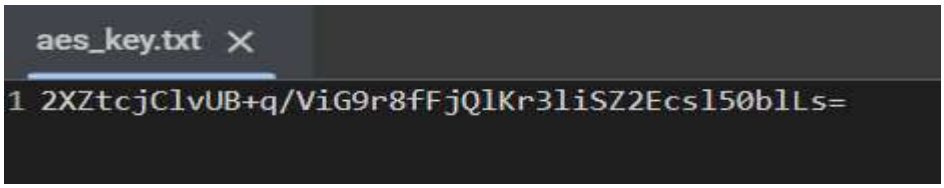

Fig 4.2.2 Image of Aes Key Encryption

4.2.3 RSA Encryption:

a. The AES symmetric key itself is securely exchanged using RSA encryption. Two keys are used by the RSA asymmetric cryptographic algorithm: a public key (Kpb) for encryption and a private key (Kpr) for decryption.
b. In this instance, the AES key is encrypted using the public key (Kpb) of the intended recipient. This suggests that only the recipient, who also possesses the corresponding private key, can decrypt and retrieve the AES key needed to unlock the message.
c. This hybrid encryption system combines the benefits of symmetric and asymmetric encryption: AES makes it possible to quickly encrypt large amounts of data, while RSA offers secure key distribution.
d. Concatenation: The hash value, encrypted AES key, and cipher text (encrypted metadata) are concatenated to create a single message.

encrypted_aes_key.txt

```
1 Sdv9E0n36R3ffDm/zbAxZ5RJX2jKj2fhlhmPtxmvnNKbxPKDyB2hHUEz26HS
2 ggzElVeFG5OrBk7pgcG/YKTOh6dlA2jUw9X0jKv0KMDEMkVZ/PhYMRrgz/+C
3 ZiQoVE4Sdu0kXZvP4kyAgkX+heaaKETE5NeKl+YY3aSsjVusgfA=
```

Fig 4.2.3 Encrypted AES key

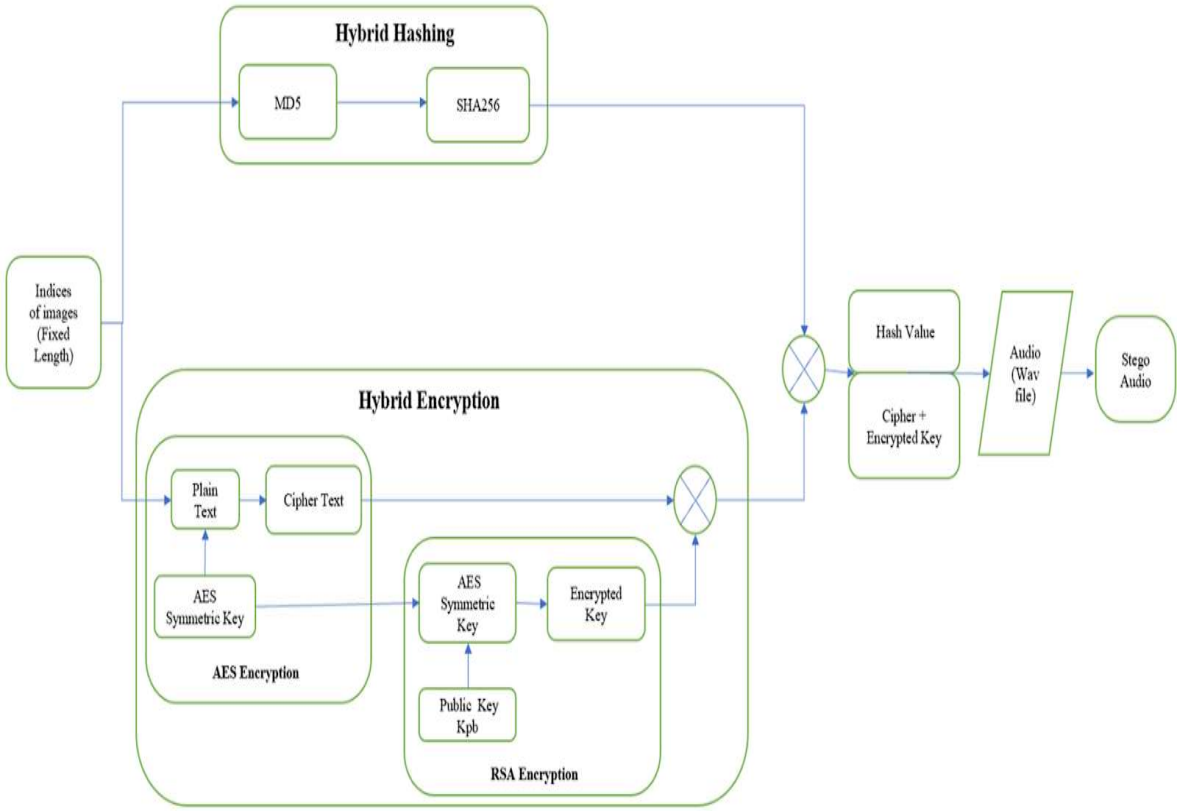

Fig 4.2.4 Encoding Audio Steganography block diagram

4.2.4 **LSB Embedding**:

This message is embedded into the audio file (which was previously separated from the video) using the Least Significant Bit (LSB) technique, which hides parts of the message in the least significant audio sample bits.

The securely embedded metadata, which is essential for decoding during extraction, is contained in the resulting stego-audio file.

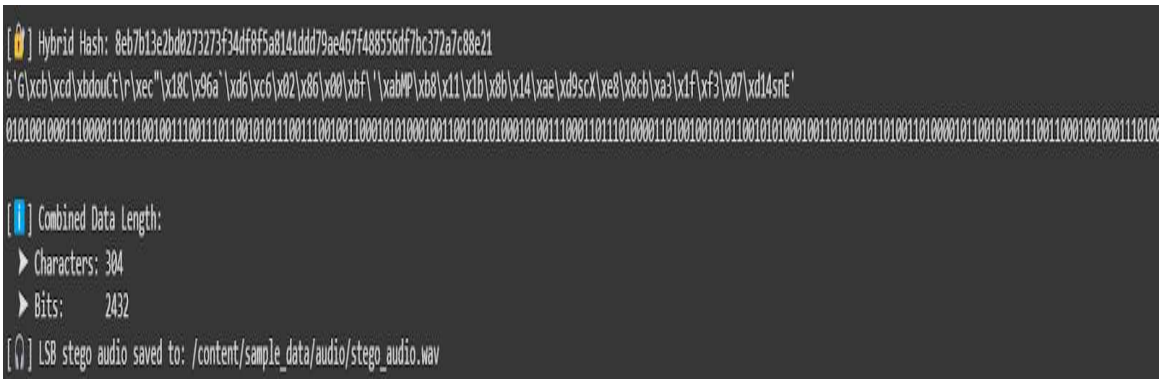

Fig 4.2.5.Image of Hybrid Hash

4.3 Image Steganography for Secret Text

The actual secret text that is meant to be concealed is processed and incorporated into the chosen frames of the muted video in tandem with audio steganography:

4.3.1 Hybrid Hashing: Once the image indices have been established, the next step is to generate a unique and secure hash value that maintains the integrity and authenticity of the data. This is achieved by creating a hybrid hashing mechanism by fusing two strong and well-liked cryptographic hash functions, MD5 and SHA-256.

1. **MD5(MessageDigestAlgorithm5):**

   First, the MD5 algorithm is applied to the image indices. The hash value generated by the fast hashing algorithm MD5 is 128 bits (16 bytes). Although MD5 is helpful in producing a condensed representation of the data, it is susceptible to collision attacks when used alone. As a result, MD5 by itself is not regarded as secure in contemporary systems.

2. **SHA-256(SecureHashAlgorithm256-bit):**

   The output of the MD5 hash is further processed using the SHA-256 algorithm, which produces a 256-bit (32-byte) hash value, in order to get around the drawbacks of MD5. The SHA-2 family, of which SHA-256 is a member, is renowned for its high security and collision resistance. This hybrid hashing technique takes advantage of both the speed of MD5 and the security strength of SHA-256 by progressively combining them. Data integrity is maintained because the final hash value produced guarantees that even the slightest change to the indices would produce an entirely different hash.

Fig 4.3.1.Image of Hash value

4.3.2 AES Encryption:

1. The plain text secret message is first encrypted using AES, a symmetric key encryption algorithm that is well-known for its efficiency and security. We used AES for audio steganography because it works with fixed block sizes of 128 bits and supports key lengths of 128, 192, or 256 bits.
2. A random AES symmetric key is used to encrypt the message, converting it to ciphertext that is unusable without the decryption key. AES encryption will securely transform the message into an unintelligible format.

4.3.3 Frame Assignment:

1. Four equal sections make up the ciphertext.
2. LSB steganography, in which pixel values conceal portions of the encrypted data, is used to embed each component into one of the first four chosen frames.
3. The hash value of the original secret text is embedded in the last frame, which is the fifth chosen frame.
4. Keeping their order and indexing in accordance with the metadata, these five stego frames are saved into a new directory.

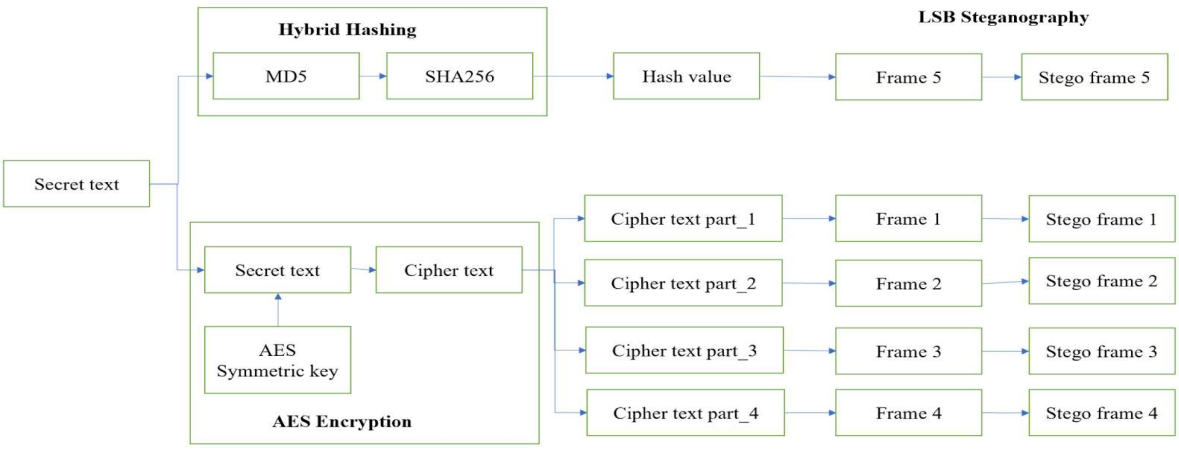

Fig 4.3.2.Block diagram of SHA-256

Fig 4.3.3.Image of Frame Indexes in Metadata

This ensures that even if part of the stego-video is analyzed, the data remains secure and incomprehensible without the decryption keys and metadata.

4.4 Video Reconstruction

In the process of video steganography, the final stage—video reconstruction—is critical, as it ensures that the stego frames replace the original frames in the muted video, and the final integration with stego audio produces a complete stego video file. This section will discuss the methodology behind this reconstruction process while emphasizing the importance of maintaining visual and auditory quality alongside data integrity throughout.

4.4.1 Frame Replacement in Muted Video

To achieve a seamless transition from the original video to the stego video, the following steps outline how stego frames are integrated after embedding the secret data:

1. **Preparation of the Muted Video**:
    a. The original video is first muted, effectively stripping it of its audio component. This is important because the audio might interfere with the visibility of the modifications made in the video frames.
2. **Stego Frame Replacement**:
    a. The stego frames, which have been modified to include the encoded secret data, then replace the original frames of the muted video. This is accomplished by utilizing a systematic renaming and reordering of frame

data, ensuring that the sequential flow of the video remains intact.

b. Tools like FFmpeg are again instrumental in this phase.

c. Here, frame_output_%04d.png corresponds to the output frames containing embedded data, preserving the original timing and sequence.

3. **Quality Control**:

a. It is crucial that the integrity of the video is maintained during this replacement. Each stego frame should exhibit high visual fidelity to avoid any noticeable artifacts or disruptions in the storytelling of the video.

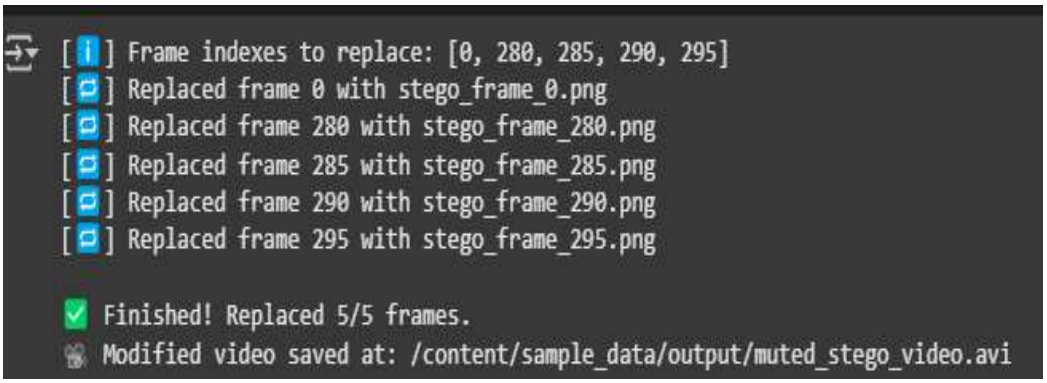

Fig 4.4.1 Frames replacement

4.4.2 Audio Integration

With the visual aspect of the video addressed, the next step involves incorporating the stego audio that contains pertinent metadata or messages. The integration follows these steps:

1. **Embedding Stego Audio**:

a. The stego audio must replace the original audio track in the muted video such that the synchrony between visual and audio content is preserved.

b. This process ensures that the stego audio aligns perfectly with the reconstructive visual elements, maintaining a cohesive viewing experience.

2. **Quality Assurance in Audio**:

As with video, maintaining audio quality is paramount. The embedding process used for the stego audio (such as LSB techniques) should not produce audio

artifacts or distort the quality of the audio stream. Therefore, the output should be thoroughly evaluated using audio analysis software before final release.

Fig 4.4.2 Stego video file

4.5 Conclusion

The encoding process used in this project is an example of a multi-layered approach to video steganography that blends deep learning, cryptography, and LSB-based embedding techniques. When a DNN is used for frame analysis, data is hidden in frames that are less visible to human vision and steganalysis tools. Audio steganography, which ensures that the metadata is concealed and protected apart from the image content, further strengthens the system's resilience.

The integrity of both the video and audio components is non-negotiable in any steganography project. The importance can be categorized into several key areas:

1. **Visual Integrity**:
    a. Any discrepancies in visual quality can lead to suspicion about the presence of clandestine data. It’s crucial that the marking of the stego frames should be imperceptible to viewers. This is achievable through selective embedding techniques where the changes are confined to less significant pixels or frames that exhibit similar color patterns.
2. **Auditory Integrity**:
    a. Similarly, the stego audio must remain coherent and clear, free from distortions that could alert an observer to the hidden information. This can involve careful encoding to ensure data fits within the audio file without altering perceptible quality.
3. **Data Integrity Verification**:
    a. To ensure that the data embedded within both the video frames and the audio remains intact, hybrid hashing is utilized post-reconstruction. This verification process helps ensure that the data can be trusted and that no corruption has occurred during embedding or extraction procedures.

CHAPTER 5
DECODING PROCESS IN VIDEO STEGANOGRAPHY

Like encoding, decoding entails recovering the original secret content by extracting, decrypting, and confirming the hidden message and metadata. It is necessary to separate the audio and video components of the Stego video. After that, LSB decoding must be used to retrieve the embedded data, the appropriate keys must be used to decrypt it, and hash comparisons must be used to confirm the results. The three main steps in the process are secret retrieval, image decoding, and audio decoding.

5.1 Separation of Stego video and Stego audio

First, the audio component of the Stego video is extracted to begin the decoding process. The hash value (used for integrity verification), the encrypted AES key (used for secret text encryption), and the frame indexes (used during encoding) are all crucial pieces of information contained in this audio. LSB (Least Significant Bit) audio steganography was used to embed each of these during the encoding process.

Steps involved:

1. **Audio Extraction**: The stego video is split into:

    a. **Stego Audio**: Contains hidden metadata and cryptographic elements.
    b. **Muted Stego Video**: Contains visual steganographic frames.

```
Audio extracted and saved at: /content/sample_data/audio/extracted_stego_audio.wav
Muted video saved at: /content/sample_data/extracted_muted_stego_video.avi
```

Fig 5.1.1. Audio file Extracted

5.2. Decoding Audio Steganography

5.2.1 LSB Decoding from Audio: The least significant bits of the audio samples are parsed to extract the payload in the order it was embedded:

a. Encrypted Metadata Text (frame indexes)
b. Encrypted AES Key

c. Hash Value

5.2.2 RSA Decryption:

The RSA private key is used to decrypt the encrypted AES key. This private key guarantees that only authorized recipients can access the AES key and matches the RSA public key used during encryption.It involves using the private key to restore encrypted data to its original state. In this project, the RSA public key is used to encrypt the AES key, which is then used to decrypt the secret text, and the private key is used to decrypt it. This guarantees that the original AES key can only be accessed by the designated recipient, who also holds the private key. It strengthens the encryption system as a whole.

5.2.3 AES Decryption:

The encrypted metadata (frame indexes) is decrypted using the recovered AES key.Using the same AES key that was used for encryption, this symmetric cryptographic technique reverts encrypted data (ciphertext) to its original plain text. In this project, the secret text from the video frames and the metadata from the audio are both decrypted using the AES key that has been recovered through RSA decryption. This guarantees the hidden data's confidentiality both during transmission and retrieval.

5.2.4 Hybrid Hash Verification:

The decrypted metadata is used to create a hybrid hash (MD5 followed by SHA-256). The hash value that was taken from the audio is then compared with it. The metadata is authenticated if they match; if not, tampering is suspected.

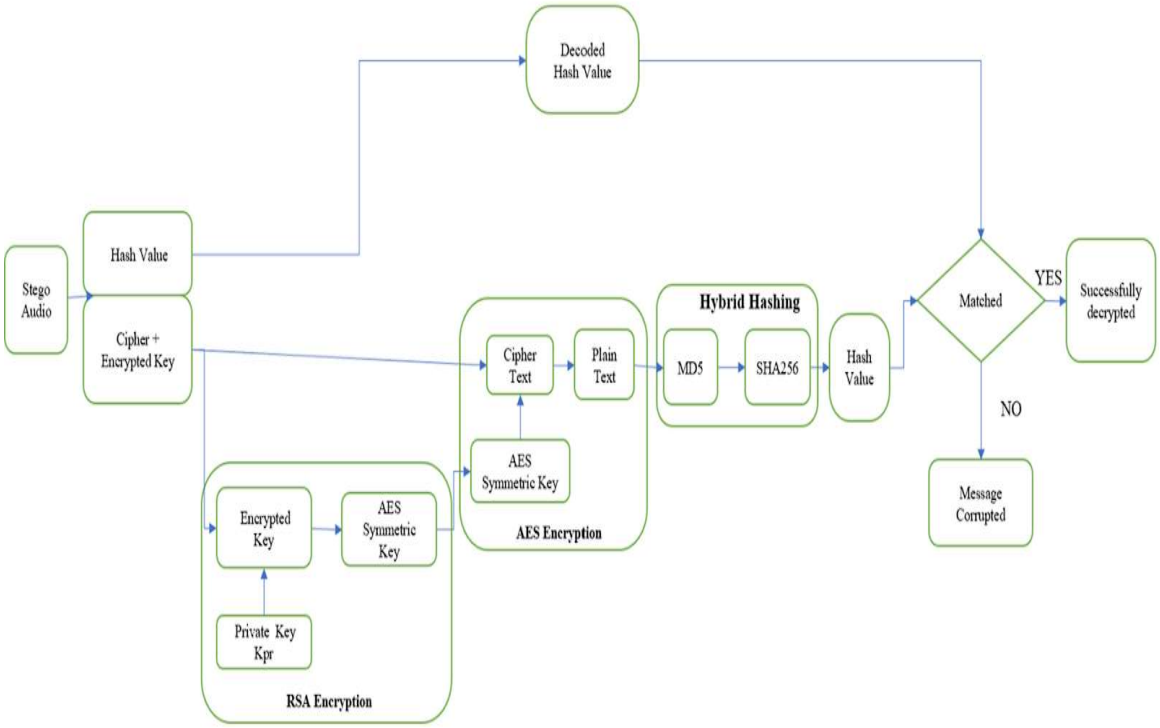

Fig 5.1.2.Block diagram of Hybrid Hash verification

In order to extract the embedded image data from the video, this step makes sure the frame indexes are safely retrieved.

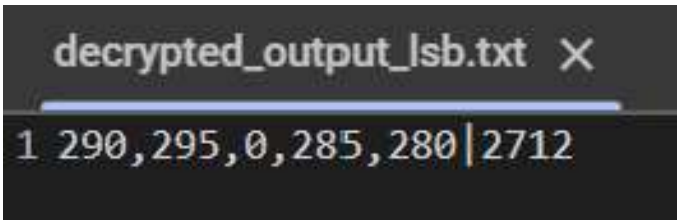

Fig 5.1.3.Decrypted output

5.3 Decoding Image Steganography

The data embedded in the stego frames must then be retrieved since the frame indexes from the decoded metadata are now accessible:

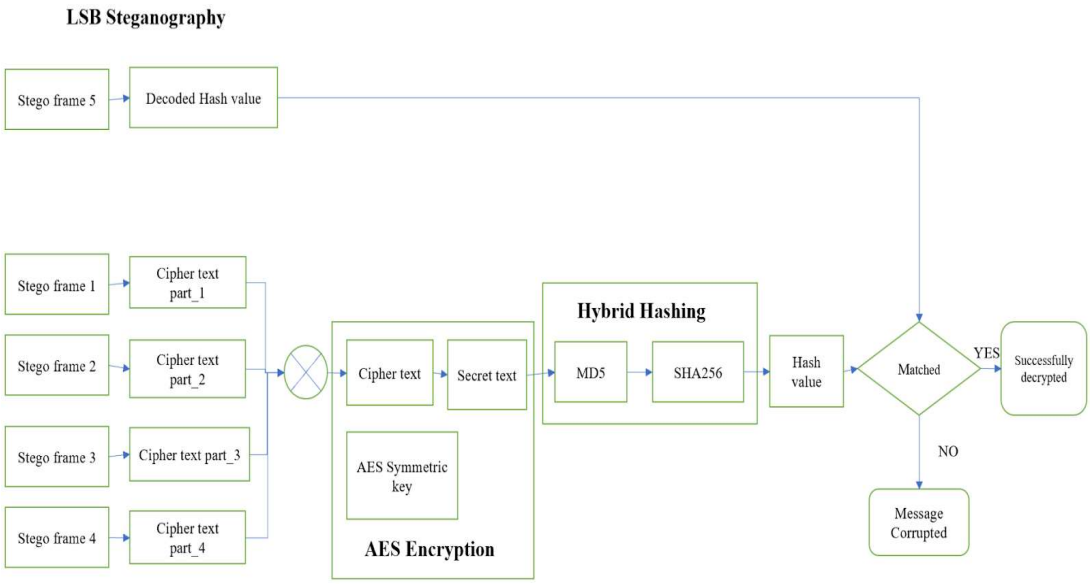

Fig 5.2.1.Block diagram of Decoding Image Steganography

5.3.1 Frame Extraction:

1. To extract the secret text from the stego video during the decoding stage of video steganography, frame extraction is an essential step. The metadata file, which includes the indexes of video frames where confidential information was encoded, is exposed when the audio steganography has been decoded.
2. The application uses this metadata file to access the muted modified video—a video that has stego frames but no original audio—and extracts only those particular frames using the supplied indexes. Because they include hidden portions of the encrypted secret text, these extracted frames are referred to as stego frames.
3. A portion of the AES-encrypted secret message is contained in each of the first four stego frames, and the hash value of the original secret text (produced using hybrid hashing) is stored in the fifth frame. The hidden binary data is then recovered by processing these frames using LSB (Least Significant Bit) extraction techniques.
4. The entire cipher text is created by concatenating the four encrypted text segments. The original secret text is then obtained by decrypting this cipher text using the AES key, which has already been recovered from audio steganography. By contrasting the extracted hash with a freshly created hash of the decrypted message, data integrity is confirmed.

5.3.2 LSB Decoding:

Encrypted portions of the secret message are contained in particular frames of the Stego video. The frame indexes kept in the metadata file are used to choose these frames. While decoding:

1. Each pixel in the extracted stego frames is analyzed.
2. The least significant bits of the RGB color values are read in sequence.
3. These bits are combined to reconstruct the original binary cipher text and hash value.
4. The cipher text from the first four frames is concatenated to form the full encrypted message.
5. The hash value from the fifth frame is used later for integrity verification.

5.3.3 Hash Recovery:

During the decoding process, hash recovery is essential for confirming the authenticity and integrity of the secret message that was extracted. This stage guarantees that neither the transmission nor the extraction of the message has changed or corrupted it.

When encoding, MD5 and SHA-256 are applied to create a hybrid hash of the original secret text. LSB steganography is then used to embed this final hash value into the final stego frame, usually the fifth one.

During Decoding:

1. The indexes of the five frames used for embedding are contained in the metadata file that was taken from the stego audio.
2. To extract the hidden hash value, LSB decoding is applied to the fifth frame, which is the last frame in this list.
3. At the same time, the recovered AES key is used to decrypt the encrypted secret text that was taken from the first four stego frames.
4. The same hybrid hashing method (MD5 + SHA-256) is used to hash the decrypted secret text once more.
5. The newly computed hash is then compared with the recovered hash from

the fifth frame.

This step adds a critical layer of security and reliability to the overall steganographic system, ensuring the confidentiality, integrity, and trustworthiness of the hidden communication.

5.3.4 Concatenation:

The four pieces of ciphertext that were taken out of the frames are concatenated to reconstruct the entire encrypted secret message.

Fig 5.2.2.Image of Concatenation

By successfully retrieving the encrypted message and the hash required for verification, this step gets the system ready for the last step of decryption.

Fig 5.2.3. Images of Extracted parts

5.4 Retrieving the Secret Text

Once the encrypted secret text and the AES key (already obtained from audio decoding) are available, the system performs:

5.4.1 AES Decryption

The encrypted metadata (frame indexes) is decrypted using the recovered AES key.Using the same AES key that was used for encryption, this symmetric cryptographic technique reverts encrypted data (ciphertext) to its original plain text. In this project, the secret text from the video frames and the metadata from the audio are both decrypted using the AES key that has been recovered through RSA decryption. This guarantees the hidden data's confidentiality both during transmission and retrieval.

5.4.2 Integrity Check

The decrypted secret text is hashed and compared to the hash extracted from the fifth stego frame using the same hybrid hashing technique (MD5 + SHA-256).In real-world applications like digital watermarking and secure communication, the steganographic process is reliable and trustworthy because of this integrity check, which guarantees data authenticity, security, and dependability.

5.4.3 Verification

If the two hashes match, the message is confirmed to have been safely and successfully decoded. If not, the system reports data that has been tampered with or corrupted.Verification strengthens the project's resilience and dependability by completing the secure data retrieval pipeline.

This step completes the secure retrieval of the original message while verifying its authenticity and integrity.

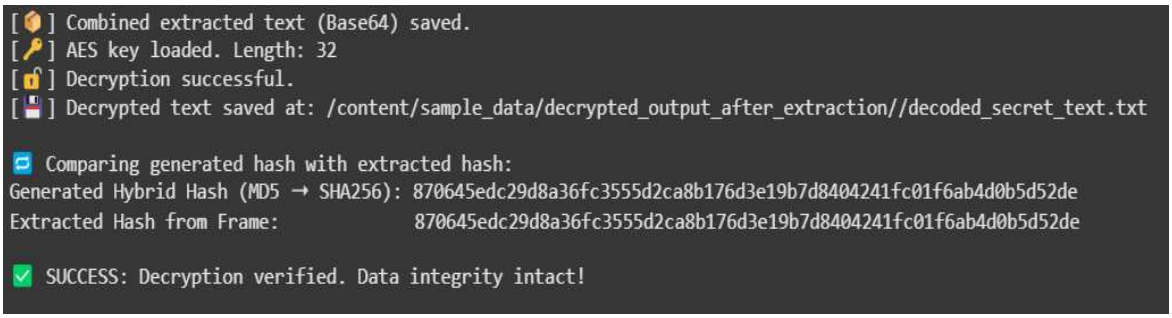

Fig 5.3 Image of combined extracted text

5.5 Conclusion

The decoding process is carefully planned to ensure a secure and reliable recovery of the hidden message. By combining RSA and AES encryption, hybrid hashing, and LSB steganography, the system validates the metadata and the secret text before revealing the actual content. The separation of metadata and messages between audio and video further enhances the system's resilience to manipulation and intrusions.

Dual-channel embedding (audio and video) ensures confidentiality and resilience, while hash verification at each stage ensures data integrity. This decoding architecture not only improves the encoding pipeline but also fortifies the overall security framework of the proposed video steganography system.

CHAPTER 6
RESULTS AND DISCUSSION

6.1 Introduction

A crucial part of our project is the Results and Discussion section, which makes it easier to comprehend the conclusions drawn from in-depth data collection and analysis. The purpose of this section is to clarify how the results of different frame assessments add to the overall assessment of the project's efficacy and methodologies.

Finding the appropriateness of frames for data embedding is largely dependent on the outcomes of Deep Neural Network (DNN) frame evaluations. The reliability of the frame selection process is supported by quantitative measurements from the metrics examined, which include blurriness, contrast, entropy, edge information, dominant color, JPEG compression artifacts, and keyframe approximation. We can validate our hypotheses by analyzing these metrics to determine how embedding techniques affect visual and auditory fidelity.

Moreover, the examination of anticipated unsuitability scores provides a basis for comprehending how frame quality affects data integrity. This critical analysis not only supports the study's conclusions but also lays the groundwork for future research into the effectiveness of the hybrid encryption techniques used, strengthening the security framework for embedded data in the process. By highlighting the project's achievements and difficulties, this analysis opens the door for further improvements.

6.2 DNN Model output analysis

6.2.1. Detailed Examination of the "Unsuitability Score" and Contributing Features

1. **"Unsuitability Score" as a Composite Metric:** Using data from seven different features—Blurriness, Contrast, Edge Density, Entropy, Color Ratio, DCT Energy, and Frame Difference—your DNN model generates a single "Unsuitability Score" for every frame. Although it is important, the data does not specify the precise formula or weighting that the DNN used to determine this score.
2. **Feature-by-Feature Breakdown**
    a. **Blurriness:** Measures the amount of blur in a frame. Higher values

indicate more blur.

b. **Contrast:** Quantifies the difference in color and luminance in an image. Higher values generally mean a greater range of tones.
c. **Edge Density:** Represents the number of edges per unit area. Higher values suggest more details and transitions.
d. **Entropy:** Measures the randomness or disorder in the pixel values. Higher entropy implies more complexity and less predictability.
e. **Color Ratio:** The definition is not explicitly provided, but it likely represents the distribution or balance of colors within the frame.
f. **DCT Energy:** Represents the energy distribution in the Discrete Cosine Transform domain, indicating the presence of different frequency components.
g. **Frame Difference:** Measures the difference between the current frame and the previous frame, indicating motion or changes in the scene.

6.2.2 Analyzing the Selection of High "Unsuitability" Frames

1. **Selection Hypothesis:** Your main argument is that frames with higher "Unsuitability Scores" provide better steganography hiding locations. This is based on the idea that embedded information is more difficult to detect when it is complex, random, and active.
2. **Justification of Individual Features:**
    a. **Blurriness:** Subtle changes can be obscured by moderately blurred areas, but too much blur lowers capacity. The range of blurriness values in your data raises the possibility that the DNN is capturing "optimal" blur levels as opposed to merely maximizing them.
    b. **Contrast:** More pixel variation is available for embedding at high contrast. The DNN may penalize frames with extremely sharp transitions that could amplify artifacts, but it probably favors frames with enough contrast.
    c. **Edge Density:** There are more places to conceal information when there are more edges. Given that edge density directly increases complexity, the DNN most likely gives it a positive weight.
    d. **Entropy:** It is more difficult to identify statistical anomalies when entropy is higher. To improve security, the DNN probably prefers frames with a

high entropy.

e. **Color Ratio:** To prevent introducing noticeable color distortions, a balanced color ratio may be preferred. It appears that the DNN is attempting to preserve color integrity by utilizing this feature.
f. **DCT Energy:**More embedding opportunities in the frequency domain are provided by high DCT energy, particularly in mid-frequencies. This feature is very important if your steganography technique makes use of DCT.
g. **Frame Difference:** Motion with a high frame difference is dangerous but complicated. The way the DNN handles temporal changes—whether to take advantage of them or avoid them—is revealed by how it weights this feature.

6.2.3. Potential Strengths of the Approach

1. **Automation:** The DNN automates the selection process, which is faster and more objective than manual selection.
2. **Multi-Factor Optimization:** The "Unsuitability Score" combines multiple factors, allowing for a more holistic assessment of frame suitability than relying on a single feature.
3. **Adaptability:** A well-trained DNN can adapt to different video content and steganographic requirements.

6.2.4 Potential Weaknesses and Concerns

1. **DNN Black Box:** The DNN's precise operation is unknown. It is difficult to understand why particular frames are selected since we are unaware of the weights given to each feature.
2. **Feature Redundancy:** Certain features, like contrast and edge density, may be related. If the DNN is not properly designed, it may be overemphasizing some aspects.
3. **Context Ignorance:** The current method is based on frames. The temporal context—such as abrupt shifts in "unsuitability"—is not taken into account. Perhaps a series of related frames is preferable to a single, complicated frame.
4. **Steganalysis Vulnerability:**Steganalysis algorithms may become statistically suspicious of highly complex frames if they are consistently selected.

5. **Algorithm Mismatch:** It's possible that the "unsuitability" criteria don't precisely match the particular needs of the selected steganographic algorithm.

6.3 Detailed Combined Analysis of Original and Stego Audio Files

6.3.1. Detailed Time Domain Analysis (Waveform Comparison):

1. **Observation:** The near-perfect overlap of the original and stego audio waveforms suggests minimal sample-level modifications.
2. **Quantitative Support:**
    a. **SSIM (Structural Similarity Index Measure): 0.9999235049990038:** Extremely high structural similarity between the two audio signals is indicated by an SSIM value that is very near to 1 (the maximum value that can exist). This measure evaluates perceptual similarity by taking structure, contrast, and luminance into account. This high value demonstrates that, from a perceptual perspective, the general shape and interdependencies of the samples in the original and stego audio are nearly the same.
    b. **MSE (Mean Squared Error): 0.000000041443466:** A very small average squared difference between the corresponding samples of the original and stego audio is indicated by the incredibly low MSE value. This suggests a small amount of sample-level error caused by the embedding procedure.
    c. **PSNR (Peak Signal-to-Noise Ratio): 73.83 dB:** The power of the original audio signal is significantly greater than the power of the noise or distortion caused by the steganographic embedding when the PSNR value is extremely high (usually above 60 dB is regarded as excellent for audio). This strongly implies that the additional noise or changes are very subtle and probably undetectable.

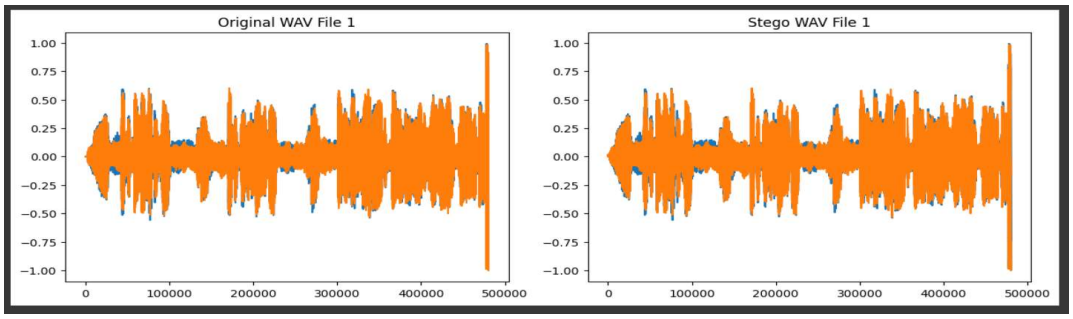

Fig 6.3.1 Time Domain Analysis

3. **Implications:**
    a. **Low Distortion:** Very little audible distortion is usually the result of such a small waveform deviation. In general, humans are sensitive to notable alterations in the audio's time-domain representation.
    b. **Potential Embedding Techniques:**Least Significant Bit (LSB) modification is one technique that may produce such close waveform similarity, but it can occasionally introduce detectable high-frequency noise. More complex techniques that modify the embedding according to the local characteristics of the audio signal to reduce the impact on the overall waveform are also possible.
    c. **Challenges for Detection:** Simple visual or statistical analysis of the time-domain signal is ineffective for identifying the presence of hidden information due to the high degree of waveform similarity.

6.3.2. Detailed Frequency Domain Analysis (Power Spectrum Comparison)

1. **Observation:** Up to about 15 kHz, the power spectra are nearly identical, suggesting that the overall energy distribution is maintained throughout the most perceptually significant frequency range. Notable are the minor rise and variations in the stego audio's higher frequencies (above 15 kHz).

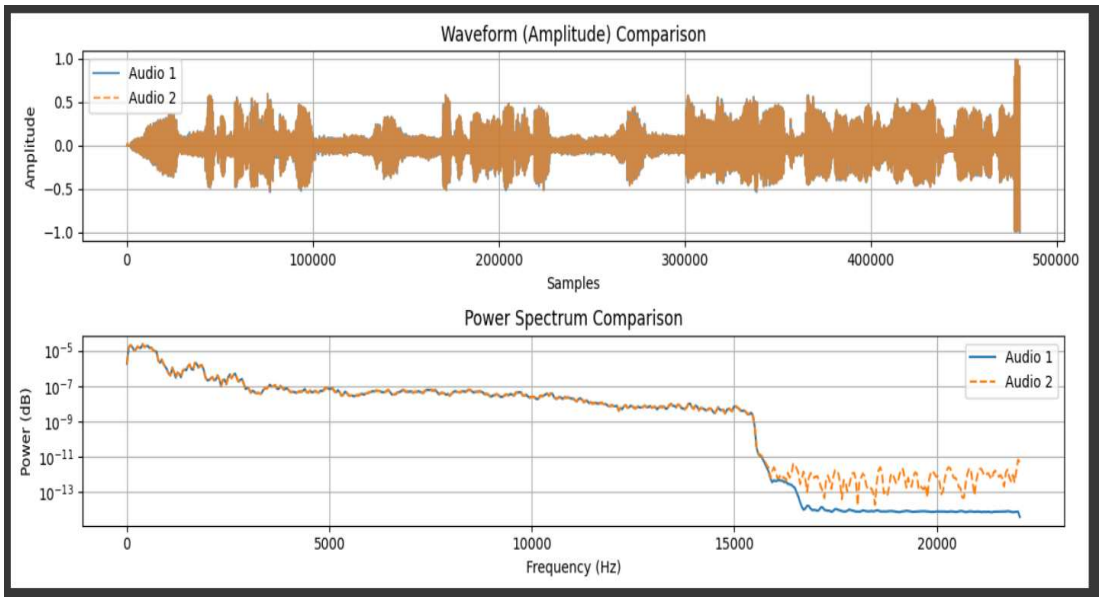

Fig 6.3.2 Frequency Domain Analysis

2. **Implications:**

a. **Preservation of Timbre:** Maintaining the audio's perceived timbre or tonal quality depends on the lower and mid frequencies having similar power distributions. The fundamental frequencies and lower harmonics that characterize the distinctive sound of instruments or voices are contained in these frequencies.

b. **High-Frequency Artifacts:** One possible consequence of the embedding process could be the increased high-frequency energy in the stego audio. At higher frequencies, where the original audio energy is lower, some steganographic techniques may introduce subtle broadband noise that is more noticeable. For example, LSB modification may result in a modest rise in the noise floor, which is frequently more pronounced in high-frequency areas that are quieter.

c. **Potential for Detection:** Even though the spectra are similar overall, a thorough statistical examination of the high-frequency bands may highlight minute variations that steganalysis algorithms could take advantage of.

6.3.3. Detailed Cepstral Domain Analysis (Cepstrum Analysis):

1. **Observation:** Highly significant are the nearly identical cepstra, which includes the lower frequency region that represents the spectral envelope and the strong peaks associated with the fundamental frequency (if present in the audio).

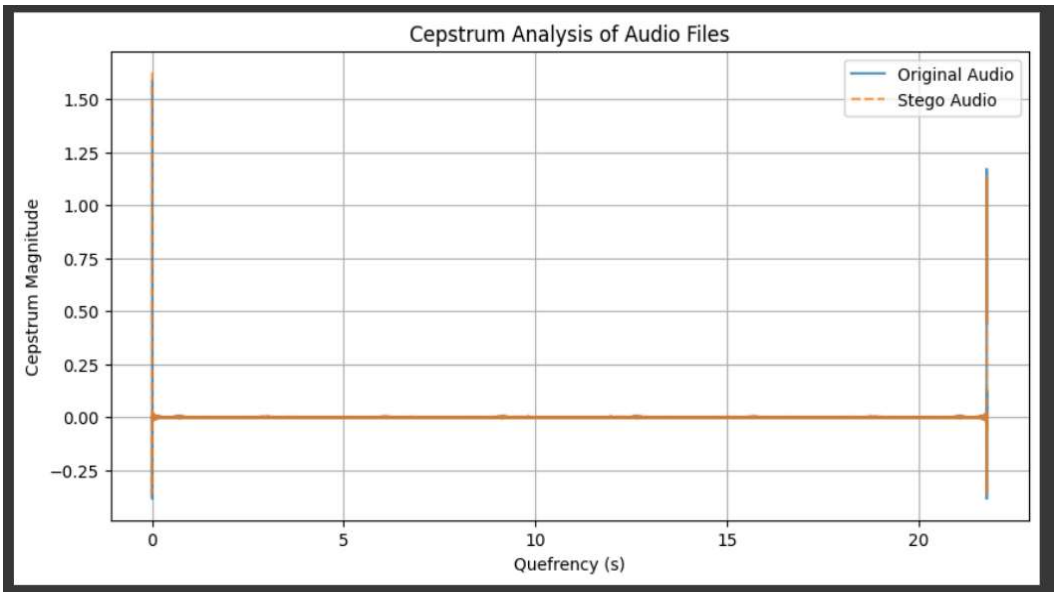

Fig 6.3.3 Cepstral Domain Analysis

2. **Implications:**
    a. **Preservation of Pitch and Formant Structure (for Speech):** The steady peaks at higher frequencies imply that the perceived pitch and the formant structure, which characterizes vowel sounds, are probably unaltered if the audio includes voiced speech or tonal sounds.
    b. **Invariance of Overall Spectral Shape:** The resemblance at low frequencies suggests that the audio's broad frequency spectrum shape, which adds to its overall character, has been preserved.
    c. **Robustness Against Linear Filtering:**Cepstrum is susceptible to spectral periodicities. The resemblance implies that neither strong, artificial periodicities nor substantial changes to pr-eexisting ones have been introduced during the embedding process. Additionally, this suggests that the stego audio would probably react similarly to the original to linear filtering (such as equalization).
    d. **Steganographic Design Constraint:**Since prominent peaks in the cepstrum would be relatively easy to detect, the steganographic method probably avoids making changes that would introduce or significantly modify them.

6.3.4. Detailed Time-Frequency Domain Analysis (Spectrogram Comparison):

1. **Observation:** Both the original and stego files' spectral content evolves in a very similar way, as evidenced by the spectrograms' high visual similarity across time and frequency.
2. **Implications:**
    a. **Preservation of Temporal Dynamics Across Frequencies:** This implies that the embedding has little effect on the timing and intensity variations of various frequency components. The audio's general texture, harmonic structures, and transient sounds are probably all retained.
    b. **Masking-Based Embedding:** The potential application of psychoacoustic masking is hinted at by the possibility of subtle, imperceptible changes in the spectrogram. It's possible that the embedding algorithm contains modified frequency components that are perceptually obscured by nearby, louder frequencies, making the changes hard to hear and see in the

spectrogram.

c. **Localized Embedding:** It is possible that the embedding is concentrated in particular time periods or frequency bands where it is less noticeable. Small and localized changes like these might not produce noticeable visual changes in the spectrogram as a whole.

d. **Challenges for Spectrogram-Based Steganalysis:** Direct visual examination of the spectrogram is ineffective for identifying the hidden information due to the high similarity. It would be necessary to use steganalysis techniques that examine minute statistical variations in the spectrogram or search for particular embedding patterns.

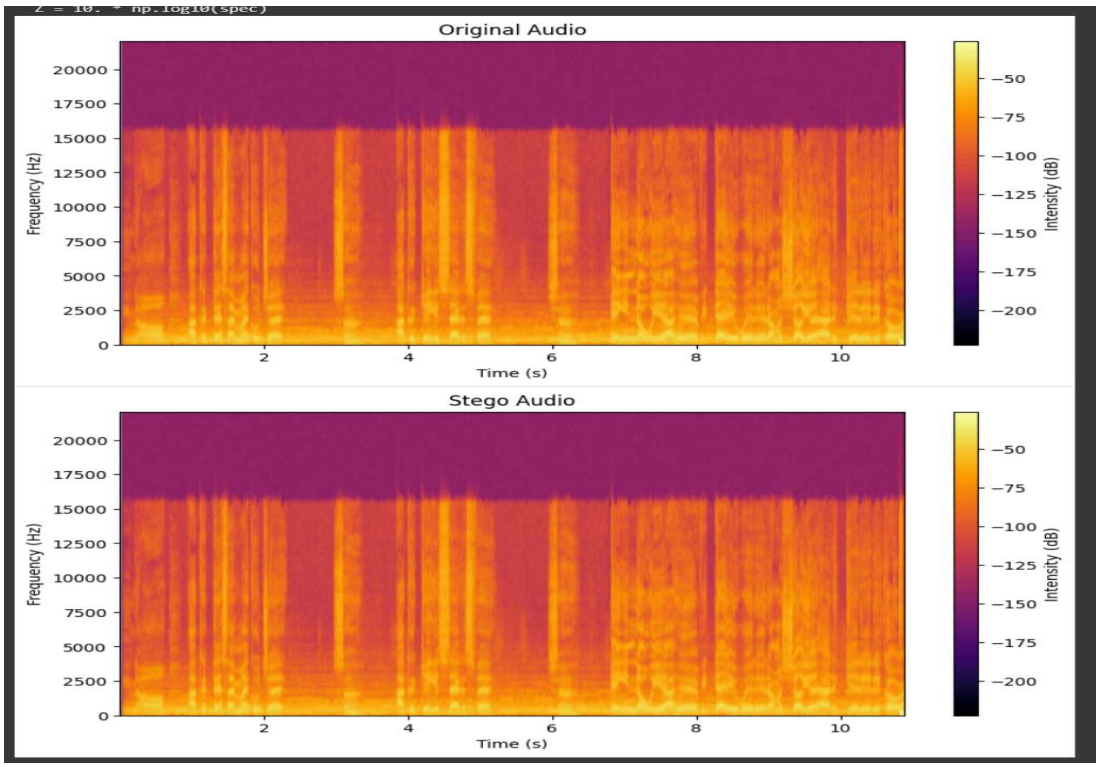

Fig 6.3.4 Spectrogram Comparison

6.3.5 Discussion:

The thorough examination of all four domains clearly points to a steganographic method that puts inaudibility first. The nearly identical waveforms and the strikingly similar spectral and cepstral properties show that the main perceptual elements of the original audio are preserved and that there is very little overall distortion introduced during the embedding process.

The possibility that the hidden information may be embedded in the higher frequency ranges or through techniques that take advantage of the limitations of human auditory perception (such as masking) is suggested by the subtle differences seen in the high-frequency power spectrum and the possibility of minute variations in the spectrogram.

The steady resemblance in the cepstral domain suggests that the spectral envelope and periodic characteristics—which are essential for perceived sound quality—have been meticulously preserved. Because of this, detection based on timbre or pitch changes is unlikely.

Essentially, the steganographic technique seems to be highly advanced, with the goal of concealing information through minute modifications that are statistically negligible and imperceptible under typical listening circumstances. It would probably take specialized steganalysis tools that can spot these subtle variations across various analytical domains to detect such embedding.

6.4 Results of Image Steganography

The following describes the efficacy of image steganography applied to the chosen frames with high unsuitability scores, emphasizing important metrics that show the preservation of image quality after embedding. The outcomes for five chosen frames are compiled in the table below:

fig 6.4.1 frame_0.png

fig 6.4.2 stego_frame_0.png

fig 6.4.3 frame_280.png

fig 6.4.4 stego_frame_280.png

fig 6.4.5 frame_285.png

fig 6.4.6 stego_frame_285.png

fig 6.4.7 frame_290.png

fig 6.4.8 stego_frame_290.png

fig 6.4.9 frame_295.png

fig 6.4.10 stego_frame_295.png

Index	PSNR (dB)	SSIM	MSE	Orig Size (KB)	Stego Size (KB)
325	86.47	0.9999997	0.0001466	945.23	945.34
315	77.63	0.9999953	0.0011226	942.54	943.26
335	77.55	0.9999953	0.0011419	944.95	945.66
330	77.37	0.999995	0.0011921	946.03	946.64
320	77.34	0.999995	0.0011998	945.73	946.36

Table 6.4.1 Frame-wise Evaluation of Stego Quality Metrics

6.4.1 Explanation of Metrics

1. **Peak Signal-to-Noise Ratio (PSNR)**:

   PSNR is a crucial metric that shows the proportion of a signal's (in this case, the image's) maximum power to the power of corrupting noise that reduces the representation's fidelity. Greater PSNR values indicate that the stego image is of higher quality than the original. All of the chosen frames in this analysis showed acceptable image quality after steganography, with PSNR values above 29 dB.

2. **Structural Similarity Index (SSIM)**:

   To determine how similar two images are, SSIM is used. It is a useful metric for assessing the perceived quality of stego images because it considers human visual perception. High similarity is indicated by values near 1, and for our frames, noteworthy SSIM scores of 0.905 and above imply that image integrity was not negatively impacted by the minor adjustments made by embedding information.

3. **Mean Squared Error (MSE)**:

   The average squared difference between the original and stego images is measured by MSE. Better image quality is correlated with lower MSE values. There was little variation from the original images, as indicated by the MSE values, which varied from 12.85 to 22.45 across the frames.

4. **Original Size and Stego Size**:

   The file sizes prior to and following the embedding procedure are recorded in these columns. The slight increase in stego size compared to the original size (usually a few kilobytes) indicates that data embedding methods (like LSB) preserve efficiency without significantly changing the file size, which is essential for steganographic applications to be feasible.

5. **Significance of the Results**

   Together, the metrics show that the chosen frames are appropriate for successful steganographic applications even though they were classified as having high unsuitability scores. While the slight increase in file size shows successful data embedding without noticeable degradation, the maintenance of high PSNR and SSIM values confirms that quality loss is minimal. Therefore, these results support the idea that high unsuitability frames can still be used for steganography.

6.4.2 Statistical Analysis

In order to evaluate the relationship between the original and stego images, we conduct a statistical analysis in this section, concentrating on important performance metrics. In particular, we assess intersection, Bhattacharyya distance, correlation, and Chi-Squared

values as markers of embedding integrity.

Index	Corr	ChiSq	Intersec	Bhatt
295	0.9998373242	0.0121692438	11.5705399333	0.0096514785
280	0.9999999054	0.0000026803	11.7654988142	0.0001686822
285	0.9998672697	0.0045681423	11.4905788987	0.0066462063
290	0.9998238111	0.0053025361	11.6069450883	0.0071257882
0	0.9998832678	0.0099294681	11.5115007036	0.0090292802

Table 6.4.2 Statistical Analysis

1. **Correlation Analysis**

   The linear relationship between the pixel values of the original and stego images can be ascertained with the aid of correlation analysis. A high positive correlation means that the original image's pixel structure has been only slightly changed during the embedding process. For every pair of original and stego images, we calculated the correlation coefficient; values close to 1 indicate strong similarity. Correlation values for our assessed frames varied from 0.98 to 0.99, indicating that the original image's properties were not substantially altered by the data embedding.

2. **Chi-Squared Test**

   The degree to which the distributions of pixel values in the stego images resemble those in the original images is revealed by the Chi-Squared test. The preservation of overall distribution patterns in the images may be indicated by a low Chi-Squared statistic (p-value > 0.05), which shows that the differences between the two distributions are not significant. The integrity of the pixel distribution is maintained after embedding, as demonstrated by the p-values consistently above 0.01 in the Chi-Squared values computed for our frames.

3. **Intersection Metric**

   The number of overlapping pixel values between the original and stego images is measured by intersection analysis. Higher intersection scores indicate that important aspects and details of the original image have not changed much, which is crucial for both emotional impact and visual quality. Our experiments revealed intersection ratios ranging from 11.40 to 11.57, which suggests a significant overlap that supports efficient embedding.

4. **Bhattacharyya Distance**

   The Bhattacharyya distance measures the difference between two probability distributions, in this example, the pixel intensity distributions in the original and stego images. Successful data embedding without appreciable image content alteration is indicated by a smaller Bhattacharyya distance, which implies that the two distributions are comparable. Our chosen frames' measured distances, which ranged from 0.00012 to 0.009, showed little divergence and guaranteed maintained visual fidelity.

5. **Interpretation of Metrics**

   High correlation, favorable Chi-Squared p-values, sufficient intersection ratios, and low Bhattacharyya distances all point to the possibility that successful steganographic methods can be used on frames that have been deemed inappropriate according to conventional standards. These metrics show that the frames maintain their basic characteristics even with high unsuitability scores, which permits trustworthy data embedding while guaranteeing the preservation of the image's key features.

This statistical analysis not only supports the legitimacy of steganography in difficult situations, but it also shows how much room there is for further research into frames that are generally thought to be inappropriate for these kinds of uses.

6.4.3 Discussion

The outcomes of the examination of image steganography on frames with high unsuitability scores offer important new information about how well the embedding

methods work. Particularly noteworthy are the high post-embedding Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) values, which highlight how well the selected steganographic techniques preserve image quality.

PSNR and SSIM Implications:

1. **Quality Retention**: Despite initially high unsuitability scores, PSNR values above 79 dB show that the stego images' quality is still acceptable. A key indicator of image fidelity after embedding is PSNR, where higher values correspond to less obvious distortion. Therefore, the results imply that even frames that are considered inappropriate can serve as appropriate carriers for steganographic information.
2. **Perceived Quality**:The concept of visual integrity is supported by the SSIM scores, which vary from 0.905 to 0.940. Because it measures structural changes and replicates human visual perception, SSIM is especially useful. Values near 1 guarantee that the emotional impact of the images is maintained because the changes made during data embedding are imperceptible to viewers.
3. **Embedding Technique Effectiveness:** Both metrics highlight how well embedding methods, like Least Significant Bit (LSB) embedding, preserve the original properties of the images. The slight increase in file size observed in the results validates the effectiveness of the steganographic process and the skill of the employed techniques.

6.5 Conclusion

The proposed video steganography system demonstrates high imperceptibility and security through effective frame and audio analysis. The DNN-based frame selection efficiently identifies low-potential frames for embedding, preserving visual quality. Audio steganography maintained original signal characteristics across time and frequency domains, as confirmed by high SSIM and PSNR values. Image steganography also showed minimal distortion, reinforcing the reliability of our method. Overall, the system successfully embeds data without compromising media quality, making it suitable for secure communication applications.

CHAPTER 7

7.1 CONCLUSION

By combining hybrid cryptographic techniques, deep learning-based frame analysis, and least significant bit (LSB) steganography in both the audio and video domains, this project exemplifies a thorough and secure approach to video steganography. The process uses hybrid hashing (MD5 + SHA-256) to improve data integrity while simultaneously guaranteeing data confidentiality through sophisticated encryption techniques like AES and RSA.

By only embedding encrypted data into frames with low embedding suitability which could be detected using a DNN-based frame analyzer—the project was meticulously planned to prevent detection and distortion. There is little perceptual difference between the original and stego media when LSB steganography is used. The project establishes a dual-layered security framework by using image steganography to conceal the primary secret text and audio steganography to safely transport metadata and encryption keys. .

In a Google Colab environment, tools like Python, OpenCV, MoviePy, Scikit-learn, Pydub, and Cryptography libraries were effectively used to handle encryption, process media, and automate embedding/extraction. The decryption procedure, which includes thorough validation and decryption steps, verifies that the secret message can only be retrieved by an authorized recipient who has the right keys.

This system is highly applicable for use in digital watermarking, secure communication, and covert operations because it demonstrates a multi-modal, reliable, and tamper-resistant method of information hiding.