A Proposed Text Encryption inside Video Using Harris Corner Detection and Salas20 Encryption Algorithm

Fadhil Abbas Fadhil, Farah Tawfiq Abdul Hussien, Teba Walaa Aldeen Khairi, Nikolai Safiullin

1Computer Science Department, University of Technology, Baghdad, Iraq.
2Engineering School of Information Technologies, Ural Federal University, Yekaterinburg, Russia.

Received 31/05/2023, Revised 27/10/2023, Accepted 29/10/2023, Published Online First 25/12/2023

Abstract

Text encryption in video is a vital tool in many industries where protecting sensitive information is of highest concern. Algorithms and automated detection methods commonly struggle to locate hidden text in video frames. This resistance increases the security of the concealed information because it is challenging for uninvited parties to detect the presence of encrypted text. This research suggests a novel method for text encryption inside video utilizing the Salsa20 encryption algorithm and Harris Corner Detection. The goal of this study is to increase the secrecy and security of textual information that is embedded in video content. First step is locatining important corners or spots inside the video frames, the Harris Corner Detection method is first used. These corners act as strong feature points that can be used to incorporate text. Without degrading the video's overall visual quality, the Harris Corner Detection method aids in precisely localizing the places where text may be concealed. Second, the encryption key is generated using the Salsa20 encryption technique. Salsa20 is a popular stream cipher that offers reliable encryption and effective operation. The embedded text is converted into a cipher text by XOR it with the encryption key before embedding it in the video, guaranteeing its confidentiality and protecting it from unauthorized access. Comparing the suggested method to current text encryption methods reveals a number of advantages. The Harris Corner Detection technique integrates text into video frames so that it blends in less noticeably, making it more difficult for adversaries to identify the existence of hidden information. Salsa20's use further guarantees the encrypted text's high level of security and protection, defending it against potential attacks. The proposed system is evaluated using MSE, PSNR, correlation, NPCR, UACI and entropy, these evaluation metrics provide excellent results.

Keywords: Harris corner detector, MSE, NPCR, PSNR, Salsa20, Text encryption, UACI, Video encryption.

Introduction

Text encryption within video is a flexible technology that improves data security, privacy, and confidentiality, making it a useful tool in many fields where safeguarding sensitive information is of utmost importance. Automated detection techniques and algorithms frequently have trouble finding hidden text in video frames. Due to the difficulty for unauthorized parties to identify the presence of encrypted text, this resistance improves the security of the concealed information. Additionally, it seamlessly conveys both forms of data at once by integrating textual and visual content. This can be helpful in circumstances
related to education, marketing, and communication.

There are many different approaches available in the present landscape of data protection and secure communication techniques, including encryption, steganography, watermarking, and others. The necessity for clandestine communication and data protection within the dynamic world of video material is frequently not adequately addressed by these methods, despite the fact that they have their advantages.

Many encryption techniques have been applied to secure the sensitive information from intruders. Multimedia must be encrypted as an important step before transmitted, multimedia involves images, audio and video. Video encryption plays an essential role in securing video sent/transmitted via mobile phones, pay TV, e-commerce, sending private emails and transmitting financial information.

According to earlier studies, video encryption techniques have been increasingly put under studying to enable secure picture transfer in order to safeguard photos from illegal access via communication channels. Text encryption inside video is a crucial task in the field of information security, as it provides a secure way of transmitting confidential messages.

Some of the gaps of the earlier studies involve:

1. Videos Don't Have Covert Communication: Most known methods for text encryption within videos concentrate first on independent encryption, then on video embedding. These techniques might not offer the level of hidden communication necessary for delicate situations, although being effective in specific situations. They frequently include obvious changes to video frames, which could tip off enemies to the existence of secret information.

2. Insufficient Resistance to Automated Detection: It is getting harder to keep hidden data in videos secret in the age of powerful automated detection technologies. Current methods might not be able to sufficiently withstand discovery by sophisticated algorithms and analysis tools, leaving buried information exposed.

3. Limited Data Integration: Some techniques ignore the possibility for seamless data integration in favor of merely hiding text within video frames. A crucial flaw in existing methods is the inability to integrate text into video footage while maintaining the integrity of both the visual and textual components.

4. Constraints on Capacity: Although certain steganographic techniques provide a high capacity for data storage within movies, they might not have the dexterity necessary for clandestine communication. Current methods frequently give capacity precedence over data privacy.

5. Implementing certain encryption and steganography techniques into video processing pipelines can be complicated and resource-intensive, rendering them unsuitable for real-time applications or situations with limited resources.

The suggested method, which makes use of Salsa20 encryption and Harris Corner detection, intends to fill these shortcomings by providing a special combination of covert communication, automatic detection resistance, data integration, and ease of use. This development aims to meet the rising demands of data protection and secure communication in today's multimedia-centric society by improving the secure transmission of text within movies while keeping a low profile in the visual medium.

One of the main challenges in this process is to embed the encrypted text in such a way that it is not easily visible to the naked eye and remains intact during the transmission. To address this challenge, the Harris Corner Detection algorithm is commonly used to identify specific points in an image or video that are robust to changes in illumination and viewpoint. Once these corner points are detected, they can be used as anchors for embedding the encrypted text. The video is essentially a combination of several edges, and each of the individual casings that make up a video has a set frame rate. The 25 edges are caught in one second because the average casing rate is 25. The video must be separated in order for this particular
computation to be carried out effectively and successfully.

In this particular instance, suppose that the movie lasts for five minutes, it can be estimated that there are 7500 edges in the film. These edges are an important structural hinderance for the video as well as for the video encryption process. This work focuses on protecting text within videos. In this instance, the attacker is unclear that the video contains text. As a result, the attacker won't make an effort to obtain its original text.

However, it is wasteful to use standard encryption techniques on digital images or video content. This is because encrypting sizable files takes a long time. So, one of the objectives of this paper is to speed up the encryption of digital data. Only a piece of the content is encrypted in partial encryption. Partial encryption shortens the time taken for encryption and decryption.

Harris corner detector is used to detect texts inside videos. A common method for identifying interest points on an image is the Harris corner detector. Despite the proliferation of feature detectors over the past ten years, this method is still widely used for camera calibration, picture matching, tracking, and video stabilization.

Overall, the combination of the Harris Corner Detection algorithm and the Salsa20 encryption algorithm provides a reliable and secure method for text encryption inside video. This approach can be applied in various contexts, including video conferencing, surveillance, cloud security, application of IoT and online communication. The primary contributions of this work are:

1. Reduce encryption time by using simple encryption technique.
2. The use of Harris Corner Detection helps in making the presence of encrypted data less conspicuous and resistant to automated detection
3. Encrypt only the text by detecting it with the Harris corner detector.
4. Utilize Salsa20 algorithm to generate strong encryption key.
   a. Key Security: Salsa20 is a symmetric encryption technique that is well-known and acknowledged for having excellent security features. It is applied here to produce encryption keys that will protect the secret text.
   b. Strong Key Generation Salsa20 uses a strong key generation mechanism to make sure that encryption keys are suitably random and secure. This is essential for maintaining the secrecy of the hidden text.
   c. Key Management: Maintaining the security of encrypted data requires effective key management. A dependable method of creating and managing encryption keys is offered by Salsa20.
5. Strong security and high speed thus can be achieved as shown by experimental results.

The remainder of the paper is structured as follows: In Section 2, the most current pertinent research is covered. The Harris corner detector is covered in Section 3. Section 4 presents Salsa20's Light Stream Algorithm. Section 5 provides a description of the suggested algorithm. Sections 6 and 7 are where the findings and recommendations are presented.

**Related Work**

Due to high need for encryption, several researchers have developed a wide variety of encryption techniques. There are lot of studies are made about video encryption. Reviewing the available works on video encryption reveals that not many of them are devoted to this topic. The most significant video encryption methods that contain texts are presented in the works below.

Ibrahim M K et al. presented a video encryption method that employs a stream cipher and a chaotic key generation scheme. The method has been successfully developed and carried out, and tests and analyses have demonstrated its effectiveness in terms of speed and security. The recommended
method has good security features and may be applied across a network for safe video exchange. The video encryption method in this research uses a chaotic system as one-time key generators. Chaos-producing systems, such as cat maps and logistic maps, are used to generate chaotic sequences that serve as encryption keys. The chosen video data is then encrypted with these chaotic sequences and stream ciphers. The pseudo-random and starting condition-sensitive sequences generated by the chaotic maps increase the security of the encryption technique.

A video watermarking technique that embeds the watermark in regions with considerable color fluctuation, such as corners, is presented by Nida F Hassan et al. This method aims to improve the security and robustness of the watermarking procedure. Using techniques for edge or corner detection, the watermark message is incorporated into the algorithm. In order to remove the watermark from the watermarked video frames, an extraction module is also included. The suggested video watermarking approach improves the security and resilience of the watermarking process by using the Advanced Encryption approach (AES) to encrypt the secret message before embedding it into the video frames. Attackers will have a harder difficulty deciphering or altering the secret message if the watermarking process is made more secure using encryption. The software also selects the ideal locations for the watermark, like corners and edges. These areas are regarded as the ideal for embedding since they have a diversity of colors, which ensures that the process won't conflict with the demands for transparent and uniform color distribution.

Firas A. Abdullatif et al. described a method for concealing encrypted data in photographs utilizing an encoding table and symmetric encryption method (AES algorithm). The Harris corner point approach is used to generate dynamic AES keys, and the encoding table is dynamically constructed from the cover image points. Then, with the exception of the corner points, the encrypted data is concealed in the least important piece of the cover image. The suggested approach exhibits strong embedding quality, flawless text recovery, and excellent PSNR value.

Sharif Shah et al. described a technique for efficient 3D stereo vision stabilization for various camera views. The method makes use of spanning tree algorithm, Laplace pyramid scaling, and Harris corner detection to find precise matching key locations in images and stable camera angles. The suggested method is rapid, accurate, and reliable and can scan more than 200 camera viewpoints in two seconds. It can be used to make photographs appear 3D and stabilize films. The method minimizes overall error for matching key-points and can be used right away for subsequent picture sets. It works well with cameras placed at random and can be improved even more by reducing image resolution to speed up processing. The 3D view reconstruction method is designed to be applied with swift processing and exact stabilization.

Revanna C R et al. proposed a technique for using variance-based quad tree decomposition to partially encrypt document images. The technique uses chaotic maps to encrypt only the significant blocks in the image after identifying their significance based on variance. The essay compares the efficacy of partial and full picture encryption and demonstrates that the suggested strategy gives superior security against a variety of threats. The recommended encryption technique provides a variety of plain images for small changes in the key, has a large key space, and is resistant to statistical attacks. When tested on several document picture kinds, the approach yields positive results.

Cheng S et al. recommend a selective video encryption solution that integrates the H.264/AVC encoding technology with encryption algorithms in order to secure video data. It uses a dynamic key generated by a PRNG along with AES in its cipher feedback mode. The technique also incorporates a 4-D hyperchaotic system to boost security. The proposed selective video encryption solution merges video coding technology with encryption algorithms by using a 4-D hyperchaotic system to encrypt significant semantic features (such as IPM, MVD, residual coefficient, and delta_QP) in each slice of the video. Selective video encryption based on video encoding and the 4-D hyperchaotic system for small
amounts of encrypted data with sufficient security, or selective video encryption based on video characteristic encoding for large amounts of encrypted data with sufficient real-time performance, are the two options available to users of this method. The combination of both video coding technology and encryption algorithms ensures the security and efficacy of the selective video encryption system.

Convolutional neural networks (CNN) are suggested by Abdulmunem IA et al. as a successful method for message obfuscation in video files. The model is made up of three parts: the prepare network, conceal network, and disclose network. After being trained on a huge dataset, the model achieves a fair level of security and embedding capacity. Performance metrics for the model include mean squared error and KL divergence. The proposed video steganography technology employs convolutional neural networks (CNN) to encrypt and decrypt video data while concealing messages. The CNN is trained to hide movies or images within other videos by deciding which portions of the cover image should be redundant and which areas can have additional pixels covered. The CNNs are designed to work in pairs and are simultaneously taught to develop the concealment and disclosure operations. It is difficult for someone without access to the weights to ascertain how the network hides the data because the weights and architecture of the CNN are random. As a result, the steganographic procedure is safer. In order to further increase security, block-shuffling is also used as an encryption technique. Additionally, picture enhancing techniques are used to improve the output's quality. Because the proposed model was trained using random RGB images from the ImageNet dataset, it has a high embedding capacity and a high security level.

Alawi A R et al. described a video encryption approach that combines the ChaCha encryption algorithm with the FAST operator in order to quickly and effectively encrypt digital videos. The approach speeds up encryption by encrypting sensitive data and evaluates performance using a range of quality indicators. The suggested video encryption methodology combines the ChaCha encryption algorithm with the FAST operator by using ChaCha as the primary encryption method and FAST as a preprocessing step to find the key spots to be encrypted. The video frames are encrypted by XORing them with a key produced by the ChaCha algorithm. The feature detection operator FAST is used to locate corners in the video frames. These perspectives are seen as potential encryption weak points. The FAST operator creates a number of essential components that change the video. The candidate points are determined by applying a segment test to each pixel in the frame and choosing a base for the computation from a circle of sixteen pixels centered on the corner candidate pixel. By combining the ChaCha encryption algorithm with the rapid operator to create quick encryption results while selectively encrypting important locations in the video frames, the proposed method decreases the encryption time and undesired interference in the transmitted films.

In order to determine which encryption method has the greatest entropy and accuracy for video frames, Khudair ET al. compare the RSA and CAST-128 encryption methods. The results demonstrate that even when dealing with distorted or ambiguous image pixels, CAST-128 surpasses RSA in terms of entropy. In the report, data encryption's importance for video transmission and storage is also stressed. Encryption includes reshaping the data into a unique shape or symbol using a secret key or password. This ensures that even if the video is intercepted or seen by unauthorized people, its content won't be interpreted or changed. According to these results, the CAST-128 encryption technique generates encrypted data that is more complicated and unpredictable, making it a more precise technique for video encryption in terms of entropy.

**Harris Corner Detector**

Major use of the Harris approach is to use intensity fluctuation in a local neighborhood to identify points. When compared to windows adjusted in any direction, the area around the feature should have a significant intensity change immediately. The autocorrelation function may be used to describe this concept as follows: make the image a scalar function $I: \Omega \rightarrow \mathbb{R}$ and $h$ surrounds any point in the domain by a modest increment, $x \in \Omega$. Corners are the locations $x$ that maximize the subsequent functional for small shifts $h$. 

---

**References**

1. Alawi A R et al. (2023).
4. ChaCha encryption algorithm.
AC(h) = \sum w(x) (I(x+h)− I(x))^2, \quad \text{...... 1}

i.e., the widest variety possible. The support area, which is commonly defined as a rectangle or Gaussian function, may be chosen by the function w(x). The equation can be linearized using Taylor expansions \( I(x+h) \rightarrow I(x) + \nabla I(x)^T h \), so that Eq 1’s right part becomes:

\[
AC(h) \approx \sum w(x)(\nabla I(x)\nabla I(x)^T)_h \\
\sum w(x)(h^T \nabla I(x)\nabla I(x)^T)_h \\
\sum w(x)(h^T h)_h \\
\sum w(x)(h^T h)_h \\
\sum w(x)(h^T h)_h \\
\sum w(x)(h^T h)_h \\
\sum w(x)(h^T h)_h \\
\sum w(x)(h^T h)_h \\
\sum w(x)(h^T h)_h \\
\sum w(x)(h^T h)_h
\]

The image gradation is determined by the autocorrelation matrix or structure tensor, which is supplied by Eq 3:

\[
M = \sum w(x)(\nabla I(x)\nabla I(x)^T) = \\
\left( \begin{array}{cc}
\sum w(x)x^2 & \sum w(x)xy \\
\sum w(x)yx & \sum w(x)y^2 \\
\end{array} \right) \\
\sum w(x)x^2 & \sum w(x)xy \\
\sum w(x)yx & \sum w(x)y^2 \\
\sum w(x)x^2 & \sum w(x)xy \\
\sum w(x)yx & \sum w(x)y^2 \\
\sum w(x)x^2 & \sum w(x)xy \\
\sum w(x)yx & \sum w(x)y^2
\right) \\
\sum w(x)x^2 & \sum w(x)xy \\
\sum w(x)yx & \sum w(x)y^2
\]

The examination of this matrix yields the maximum of Eq 2. The direction of the biggest intensity fluctuation is corresponding to M’s largest eigenvalue, whereas the orthogonal intensity variation is corresponding to M’s second largest eigenvalue. Algorithm 1 could be used to apply the Harris Technique. To minimize aliasing and picture noise, the image is initially convoluted with a low standard deviation Gaussian function.

Algorithm 1: Harris Corner Detection

input: index, ms, k, σi , τ, σi, str, cells, σi, N, sub_of_pixel
output: The Corners of Harris

Gaussian (index, σi) → index' gradient (index') → (Index_x, Index_y)

(\text{Index}_x, \text{Index}_y, \sigma_i) → (X', Y', Z') // Autocorrelation computation

(X', Y', Z', ms, k) → Res // compute corner response, where ms = measure

(Res, τ, 2σi) → corners // find corners using NOT(maximum suppression)

(corners, str, cells, N) // select output corners, where str= strategy

if sub_of_pixel

(R, corners) // accuracy of sub_of_pixel

The image’s gradient is used to compute the autocorrelation matrix, and its eigenvalues are utilized to determine if any of the preceding circumstances apply. To eliminate regions with low values in this function, a threshold must be specified. This threshold relates to the amount of noise in the pictures and depends on the corner strength function being utilized. The points in the final two phases can be chosen in a variety of ways, and quadratic interpolation can increase the accuracy of the corners 21.

Light Stream Algorithm

Different information formats may now be legally accessed by an authentic user thanks to encryption techniques 22. Any system that needs high-speed encryption, like SSL 23 and WEP 24, can benefit from effective stream cipher implementations. The goal of the ECRYPT (eSTREAM) 25 stream cipher project is to find novel stream ciphers that enable fast and safe encryption. A differentiating attack 26 and a Crossword Puzzle Attack 27, one of the distinctive attack’s variants, both succeeded in breaking this stream cipher. Two examples of light stream algorithms are Salsa and Cha-Cha. The Salsa20 stream cipher works on 32-bit words, accepts a 256-bit key as inputs k= (k0, k1, ..., k7) or a 128-bit key as input k= (k0, k1, ..., k3), and outputs a series of 512-bit key stream blocks. By XORing the key, nonce, and block number, Salsa20 encrypts a 64-byte block of plaintext. The Algorithm2 defines Salsa20 steps:

Algorithm 2: Salsa_20 stream cipher

input: initial array F, rounds ∈ N
output: Out=F+Rrounds

T‘←F

Rounds_len= rounds/2

While index < Rounds_len

//3-6: Column round

(T0, T1, T2, T3) ← SalsaQRfunction (f0, f1, f2, f3)

(T4, T5, T6, T7) ← SalsaQRfunction (f4, f5, f6, f7)

(T8, T9, T10, T11) ← SalsaQRfunction (f8, f9, f10, f11)

(T12, T13, T14, T15) ← SalsaQRfunction (f12, f13, f14, f15)

//7-10: row round

(T0, T4, T8, T12) ← SalsaQRfunction (f0, f4, f8, f12)

(T2, T6, T10, T14) ← SalsaQRfunction (f2, f6, f10, f14)

(T1, T5, T9, T13) ← SalsaQRfunction (f1, f5, f9, f13)

Out= X+T'

return (Out)

Salsa20(x)≡ x+( rowround ( columnround (x)))^R

......4

While there are R double rounds, there are only r ciphering rounds (r=2R). Then, as seen below, each block with 64-byte is XORed with the corresponding block of 64-byte from the mere image 28, 29.
Materials and Methods

One of the things that was previously discussed is that the appearance of some important texts in the video clips may be confidential and undesirable to show them to the public. A new algorithm that selects texts in the video to be encoded has been proposed. This suggested algorithm's primary function is to encrypt text pictures that are taken directly from the frames of the input video. Of course, the algorithm's input is a video clip (secret texts are expected in it). The partly encrypted video is handled in the following diagram, Fig 1.

![Figure 1. The proposed system diagram.](image)

Algorithm 3 describes the proposed work in the process of detecting and encoding text in video:

**Algorithm 3**: Proposed algorithm

**input**: Digital video file

**output**: Encrypted Video

**Step_1.** Key Generation:
- Generate key vector using Salsa20 encryption algorithm (key stream algorithm)

**Step_2.** Split the input video into frames $i$

**Step_3.** For each frame $i$ do the following:
- Apply Harris corner detector
- Encryption the frame using:
  - Encrypted frame $i = Frame(i) \oplus Key(i)$

**End for**

**Step_4.** Encrypted video = Collect encrypted frames.

**Step_5.** Return (Encrypted_video).
The following explain points the key steps involved in algorithm 3:

1. Harris Corner Detection: This computer vision technique locates distinguishing elements or locations inside an image or video frame. When the image or video is transformed, such as by rotation, scaling, or translation, these corner points are chosen because they are less likely to change their positions. To locate acceptable areas within video frames where text can be hidden in this situation, Harris Corner Detection is used.

2. Salsa20 Encryption Algorithm For Key Generation: The symmetric encryption algorithm Salsa20 is renowned for being quick and secure. On the basis of a secret key and a nonce (a one-time use only unique number), it generates pseudorandom keystreams. Salsa20 is used in this suggested method to both encrypt the content and create the encryption key. A secret key and a nonce must be provided to Salsa20 as inputs during the key generation process. Next, text is encrypted and decrypted using the generated key.

Text Encrypted Inside Video: The text that has to be hidden is encrypted using the encryption key that has been generated. Then, the encrypted text is incorporated into the video frames. In order to insert the encrypted text in the relevant corners or points within the frames, Harris Corner Detection is used. To conceal the encrypted text, embedding entails changing the pixel values or characteristics at the chosen corners.

A recipient needs to have the secret key used for encryption and key creation in order to access the encrypted text. The recipient can decrypt the embedded text and view the original message by using the same Salsa20 method with the key and nonce.

Using Harris Corner Detection, the procedure normally entails choosing video frames or images that are appropriate for text embedding. The text is then hidden at these spots by using them as corners. To create safe encryption keys, Salsa20 is utilized. These keys are then used to encrypt the text before it is embedded within the video frames. The secret text from the video frames is extracted and decrypted using the same encryption keys during decryption.

Text Encryption is used for partial text encryption and were incredibly impressed with the results. The proposed technique is robust and secure, and it was tested extensively using various samples of colored images extracted from the input video. It's a great way to keep data safe and secure, while also ensuring that it's accessible when you need it; the encryption steps are described as follows.

Text Detection

To detect the texts in the video, the Harris corner detector is used, whose method of operation has been explained in the third part of this paper. Naturally, there are several mathematical operations calculated in the background of this procedure. However, it could be a little daunting to try to explain everything. Therefore, the first three phases outline the fundamental principles of how it operates:

Step_1. The window with the greatest intensity fluctuation when moved in both the X and Y axes is found by sliding a fixed-size window over the picture.

Step_2. A score R is calculated for each window that is discovered.

Step_3. Then, by adding a threshold to this score, the required corners are chosen. R is often seen as being small when an area is flat. If R is little, the region is regarded as a corner, and if it is large, it is seen as an edge.

Figs 2, 3 and 4 represent the results of using Harris corner detector.
Figure 2. Harris corner detector for a book (A) Text Area, (B) Text Detector

Figure 3. Harris corner detector for exam paper (A) Text Area, (B) Text Detector

Figure 4. Harris corner detector for chat window (A) Text Area, (B) Text Detector

Keystream Generation
Since randomness tests are needed to confirm the produced key, it is well known in cryptography that the key, the technique of its production, and the degree of its unpredictability are crucial factors in determining how strong the algorithm is. Salsal20 encryption algorithm was used in the key generation process.

Text Encryption in Images
After detecting the edges that are represented by texts in each frame, these points are determined in order to be encrypted using the XOR function with the keys, as shown in Figs 5, 6, and 7.
Results and Discussion

Encryption is the finest method of security when it comes to digital video content protection. Thus, there have been two methods for encrypting digital video files: full encryption or partial encryption. The key was crucial to the encryption process since it improved security, strengthened the algorithm, and made it more difficult for attackers. Using the proposed algorithm, three samples of videos were used to be adopted in finding the results of the measurements. Table 1 explains the results of encryption and decryption for the three samples which were utilized in the research findings. Both in terms of encryption process and time, the recommended encryption algorithms have produced positive results. Three videos of varying sizes are used to demonstrate these values in Table 1.

<table>
<thead>
<tr>
<th>File name</th>
<th>File size</th>
<th>File type</th>
<th>No. Of frames</th>
<th>Frame sequence</th>
<th>Encryption time</th>
<th>Average Encryption time</th>
<th>Decryption time</th>
<th>Average Decryption time</th>
</tr>
</thead>
<tbody>
<tr>
<td>The book</td>
<td>2.09 MB</td>
<td>MP4</td>
<td>132</td>
<td>1</td>
<td>2.9588258</td>
<td>2.1961140</td>
<td>3.3775222</td>
<td>2.7678514</td>
</tr>
<tr>
<td></td>
<td>(2,195,456 bytes)</td>
<td>File</td>
<td>2</td>
<td>2.449038</td>
<td></td>
<td></td>
<td>3.0947067</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2.9378225</td>
<td></td>
<td>3.4764475</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>3.2716002</td>
<td></td>
<td>4.8924071</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>3.4774487</td>
<td></td>
<td>2.8778858</td>
<td></td>
</tr>
<tr>
<td></td>
<td>132</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.7389886</td>
<td></td>
</tr>
<tr>
<td>Exam paper</td>
<td>5.81 MB</td>
<td>MP4</td>
<td>386</td>
<td>1</td>
<td>2.6389703</td>
<td>2.4924267</td>
<td>2.9037544</td>
<td>2.6627626</td>
</tr>
<tr>
<td></td>
<td>(6,094,848 bytes)</td>
<td>File</td>
<td>2</td>
<td>2.6053428</td>
<td></td>
<td></td>
<td>3.5778110</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2.4282143</td>
<td></td>
<td>2.1929590</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>2.8431668</td>
<td></td>
<td>2.1455924</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>2.1439390</td>
<td></td>
<td>2.1566720</td>
<td></td>
</tr>
<tr>
<td></td>
<td>386</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.1986498</td>
<td></td>
</tr>
<tr>
<td>Chat</td>
<td>704 KB</td>
<td>MP4</td>
<td>102</td>
<td>1</td>
<td>2.5597155</td>
<td>2.3897914</td>
<td>2.5924718</td>
<td>2.5482144</td>
</tr>
<tr>
<td></td>
<td>(720,896 bytes)</td>
<td>File</td>
<td>2</td>
<td>2.7031121</td>
<td></td>
<td></td>
<td>2.7051579</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2.7308461</td>
<td></td>
<td>2.9893336</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>2.7349669</td>
<td></td>
<td>2.5461242</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>2.6140787</td>
<td></td>
<td>2.6170725</td>
<td></td>
</tr>
<tr>
<td></td>
<td>102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.3897900</td>
<td></td>
</tr>
</tbody>
</table>
In order to evaluate the strength of video encryption, human examination is not the only method, so other measurement techniques were relied upon to assess the level of encryption more objectively and accurately. There are many measures of quality factors, and the most important of these factors that were used to measure the proposed method (UACI, PSNR, NPCR, MSE, correlation and entropy)\textsuperscript{24, 30}, this is shown in Table 2.

Table 2. Fidelity Criteria

<table>
<thead>
<tr>
<th>Video File name</th>
<th>Video File size</th>
<th>Criteria</th>
<th>Encryption with Salsa20</th>
<th>Encryption with Salsa20 and Harris Corner Detector</th>
</tr>
</thead>
<tbody>
<tr>
<td>The book</td>
<td>2.09 MB (2,195,456 bytes)</td>
<td>MSE: 26328.318</td>
<td>41.3324</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PSNR: 3.926</td>
<td>32.2900</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correlation: -0.0018</td>
<td>0.7721</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPCR: 99.60</td>
<td>45.1161</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UACI: 27.77</td>
<td>12.0516</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Entropy: 7.2571 Encrypt</td>
<td>7.3523 Encrypt</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.7065 Decryp</td>
<td>7.1666 Decryp</td>
</tr>
<tr>
<td>Exam paper</td>
<td>5.81 MB (6,094,848 bytes)</td>
<td>MSE: 15744.076</td>
<td>13.6844</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PSNR: 6.159</td>
<td>37.4598</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correlation: -0.0013</td>
<td>0.8299</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPCR: 99.47</td>
<td>16.4066</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UACI: 20.19</td>
<td>4.5609</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Entropy: 0207.7 Encrypt</td>
<td>6.2913 Encrypt</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>020700 Decryp</td>
<td></td>
</tr>
<tr>
<td>Chat</td>
<td>704 KB (720,896 bytes)</td>
<td>MSE: 26983.749</td>
<td>57.0724</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PSNR: 3.819</td>
<td>30.5854</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correlation: -0.0004</td>
<td>0.2522</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPCR: 99.55</td>
<td>61.1204</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UACI: 29.21</td>
<td>19.6976</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Entropy: 0207.9 Encrypt</td>
<td>6.8072 Encrypt</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>020707 Decryp</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 2, the MSE values are high, and the PSNR obtained values are low in the above table, proving that the suggested algorithm was effective and resistant to attacks. The correlation values in the suggested algorithm were excellent. The NPCR and UACI numbers demonstrated the algorithm's high resistance to differential attacks. As it is known that the ideal value of entropy is 8\textsuperscript{8}, and whenever the result is close to this value, the results are good, also note that the results of the proposed algorithm are very good due to its closeness to the ideal value, demonstrating that the suggested encryption technique was effective in distributing the encrypted frame's pixels at random.

Comparison Analysis

A comparison analysis according to storage capacity, complexity, speed, randomness and entropy is performed between the related works and the proposed method. It is summarized in Table 3.
### Table 3. Comparison analysis

<table>
<thead>
<tr>
<th>Paper</th>
<th>Storage capacity</th>
<th>Complexity</th>
<th>Speed</th>
<th>Randomness</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Moderate</td>
<td>High</td>
<td>vary depending on the</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>algorithm and hardware used.</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Depends on the frames</td>
<td>Moderate</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>number</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>17</td>
<td>High</td>
<td>Depends on</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>computer vision, sensor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>fusion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Moderate</td>
<td>High</td>
<td>vary depending on</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>image size, hardware</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Low</td>
<td>Moderate</td>
<td>low</td>
<td>Not considered</td>
</tr>
<tr>
<td>20</td>
<td>depends on the amount of</td>
<td>Moderate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>data to be hidden, the</td>
<td></td>
<td>Depends on</td>
<td></td>
</tr>
<tr>
<td></td>
<td>chosen steganography</td>
<td></td>
<td>steganography method, ability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>technique</td>
<td></td>
<td>to disguise the hidden</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>information</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Low</td>
<td>Moderate</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>High</td>
<td>Moderate</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>The proposed method</td>
<td>Depends on the text length</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Limitations and Future Works

The suggested approach shines at data integration and covert communication, but it has text capacity restrictions. While offering more variety and capacity, separate encryption and video embedding may also be less stealthy and require a more involved approach. The decision should be based on the trade-offs between robustness, robustness, text capacity, covert communication, and implementation complexity. Despite having advantages, text encryption inside a video using Harris Corner Detection and the Salsa20 encryption method has some drawbacks and difficulties:

1. **Limited Capacity:** The amount of data that can be stored when embedding text into video frames using this method is constrained. A limited number of corners are found in video frames, and each one can only accommodate a modest amount of text. This restricts how much information can be safely concealed within the movie.

2. **Lower Video Quality:** Text inclusion in video frames has the potential to lower the video quality. When pixels are changed to conceal text, there may be abnormalities or visual distortions that become apparent only upon careful scrutiny.

3. **Robustness:** The security of hidden data raises questions. The hidden text may be destroyed or corrupted by modifications to the video, such as compression, resizing, or format conversion, rendering it impossible to restore.

4. **Detection and Attacks:** Although Salsa20 encryption and Harris Corner Detection can make it difficult to find the secret text, they do not totally protect it from advanced analytic methods. There's still a chance that sophisticated algorithms and forensics tools can find any buried data.

5. **Key Management:** Maintaining the security of the hidden text requires effective key management. Data leakage or loss of access to the secret data may result from the loss or breach of the encryption keys.

Using Harris Corner Detection and the Salsa20 Encryption Algorithm, it is possible to overcome or at least lessen the drawbacks of text encryption inside a video. This requires a combination of technical, procedural, and strategic methods. Here are several strategies for overcoming these restrictions:

1. **Capacity Limitation:** Taking into account employing more sophisticated steganography methods, such as frequency domain techniques like
Discrete Cosine Transform (DCT) or Wavelet-based approaches, which may conceal bigger amounts of data.

2. Video Quality: Using better compression methods to lessen the effect that text embedding has on video quality to minimize artifacts created during text embedding, optimize the encoding and decoding operations

3. Discovery: To make discovery more difficult, using additional steganographic techniques such data fragmentation, randomization, and encryption before embedding text. Keep up with the most recent steganalysis developments and change your strategy to thwart new detection techniques.

4. Key Management: Creating a reliable key management system to create, store, and distribute encryption keys in a secure manner. Using sound, proven key management procedures to increase security, think about utilizing asymmetric encryption for key exchange.

5. Metadata: • Removing any identifying information from video files' metadata to prevent it from revealing the existence of hidden data to reduce the chance of metadata exposure, exercise caution while sharing or distributing videos with concealed data.

6. Robustness: Testing your hidden text's resistance to different video transformations and use error-correction algorithms to restore data in the event of corruption.

7. Security Assumptions: • To ensure that the encryption algorithm (such as Salsa20) is secure, it should be regularly updated and audited. Use strict access controls and oversight to safeguard encryption keys.

8. Real-world use cases: • Determining whether video steganography is the best option for your particular use case. Alternative encryption and transmission techniques can be more appropriate depending on your needs.

Conclusion

The proposal to use Harris Corner Detection and Salsa20 Encryption Algorithm for text encryption inside video is a promising idea that can enhance the security of data transmission. Through the use of Harris Corner Detection, important features within the video frames are identified and used to extract the text that needs to be encrypted. The Salsa20 Encryption Algorithm, on the other hand, provides a fast and secure encryption process that can protect the confidentiality of the data. The experimental results showed that the proposed method is effective in encrypting text inside video frames and maintaining the quality of the video. The comparison analysis demonstrated that the proposed method outperformed existing methods in terms of security and speed. The need for secure data transfer is growing as a result of ongoing technological improvements, and the suggested solution can be crucial in addressing this requirement. Overall, the proposed method of text encryption inside video using Harris Corner Detection and Salsa20 Encryption Algorithm can provide a practical and effective solution to the problem of data security in digital communication.

Authors’ Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Besides, the Figures and Images, which are not ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Technology.

Authors’ Contribution Statement

F.A. and N.S. proposed this idea, F. T. And T.W. suggested the general outline of the proposal. Then F.A. carried out the proposal, extracted the results, and discussed it with N.S., F.T. and T.W. to suggest improvements. F. T. organized and edited the paper.
References


مقترح تشفير نص داخل الفيديو باستخدام خوارزمية هاريس لاكتشاف الزوايا و خوارزمية التشفيرسالسا 20

فاضل عباس فاضل 1، فرح توفيق عبد الحسين 1، نيكولاي سفينوئن 2

1 кафедра العلوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق.
2 كلية الهندسة للكهرباء، جامعة أورال الفيدرالية، يكاترينبورغ، روسيا.

لاكتشاف الزوايا هاريس. الأطراف غير المدعوة صعبة في تحديد موقع النص المخفي في إطارات الفيديو. تزيد هذه المقاومة من أمان المعلومات المخفية. يتم تشغيل النص في الفيديو باستخدام خوارزمية التشفير Salsa20. Salsa20 هو تشفير دقيق يوفر تحكمًا فعالًا. يتم تحويل النص المخفي إلى نص تشفير بواسطة XOR باستخدام خوارزمية Salsa20. يتم تشغيل النص المخفي بواسطة خوارزمية Harris Corner Detection، والتي تعني استخدام خوارزمية هاريس لتشخيص النقط المميزة في الفيديو. يتم تشغيل النص المخفي باستخدام خوارزمية Salsa20. يتم تخزين النص المخفي في الفيديو، مما يضمن سرية وحماية النص المشفر، والدفاع عنه ضد الهجمات المحتملة. يتم تشفير النص باستخدام خوارزمية Salsa20، وتوفر مقاييس تقييم هذه النتائج ممتازة.

الكلمات المفتاحية: تشفير النص، Salsa20-PSNR، NPCR، MSE، التشفير، النص، خوارزمية هاريس، Corner Detection.