INTERNATIONAL JOURNAL OF INTELLIGENT DATA AND MACHINE LEARNING (IJIDML)

eISSN: 3087-4262

Volume. 02, Issue. 04, pp. 01-05, April 2025

ENHANCED IMAGE STEGANOGRAPHY: LSB SUBSTITUTION WITH RUN-LENGTH ENCODED SECRET DATA

Dr. Maria Gonzalez
Department of Computer Science, University of São Paulo, Brazil

Article received: 14/02/2025, Article Accepted: 21/03/2025, Article Published: 06/04/2025
DOI: https://doi.org/10.55640/ijidml-v02i04-01

ABSTRACT

Image steganography has emerged as a vital technique for secure communication by concealing sensitive information within innocuous digital media. This study proposes an enhanced image steganography method that integrates Least Significant Bit (LSB) substitution with run-length encoding (RLE) of the secret data to improve embedding efficiency and reduce detectability. By applying run-length encoding prior to embedding, the secret message is compressed, enabling a greater volume of information to be hidden within the cover image while maintaining minimal perceptual distortion. The proposed approach adaptively selects embedding regions based on local image characteristics to further increase imperceptibility and robustness against steganalysis. Experimental results demonstrate that the method achieves higher peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) compared to conventional LSB substitution techniques without compression. This research highlights the potential of combining data compression and adaptive embedding strategies to advance the state of image steganography, offering a practical solution for secure data hiding in modern digital communication environments.

KEYWORDS

Image steganography, LSB substitution, Run-length encoding, Data hiding, Information security, Steganalysis resistance, Image processing, Compression techniques, Digital watermarking, Secure communication.

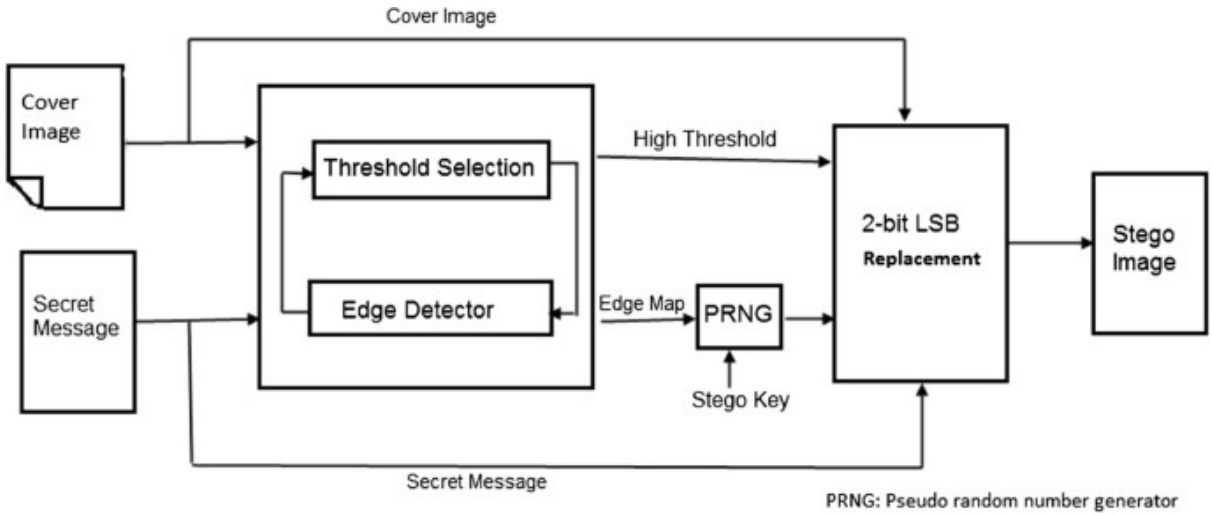
INTRODUCTION

In the digital age, the secure transmission and storage of sensitive information have become paramount. Cryptography, which scrambles data to render it unreadable without a key, is a primary method for ensuring confidentiality. However, cryptography makes the presence of hidden information obvious, which can attract unwanted attention. Steganography, on the other hand, aims to conceal the very existence of communication by embedding secret data within innocent-looking cover media, such as images, audio, or video [1]. The goal of steganography is to achieve high imperceptibility (the hidden message should not be detectable by human perception or statistical analysis) and robustness (the hidden message should resist various attacks).

One of the simplest and most widely used steganographic techniques is Least Significant Bit (LSB) substitution. This method involves replacing the least significant bits of the cover medium's pixel values with the bits of the secret message. While conceptually straightforward and easy to implement, traditional LSB substitution methods often face challenges regarding imperceptibility and security. Replacing a large number of LSBs can introduce noticeable distortion in the cover image, making the presence of the hidden data detectable through visual inspection or statistical analysis (steganalysis). Moreover, direct sequential LSB replacement can be vulnerable to simple attacks if the embedding pattern is predictable [2].

To address these limitations, various advancements have been proposed, including techniques based on Multi-Resolution Analysis (MRA) and coordinate conversion [3, 4, 5, 6, 7]. The core idea is to embed secret information in transform domains or by modifying embedding strategies to enhance invisibility and security. A critical aspect of improving steganographic performance is reducing the amount of secret data that

needs to be embedded, thereby minimizing the modification to the cover medium. Data compression techniques play a vital role in this regard. Run-length encoding (RLE) is a lossless data compression method that efficiently represents sequences of identical data values as a single data value and its count. This is particularly effective for binary images or images with large uniform areas, which often characterize secret images (e.g., text, logos, or simplified visual information).



This article proposes an enhanced LSB-based data hiding method that leverages run-length encoding for the secret image. The primary objective is to improve the imperceptibility and, potentially, the security of the embedded secret image by significantly reducing the volume of data to be hidden. By applying run-length coding to the secret image, the method minimizes the number of LSB modifications required in the cover image. Furthermore, the embedding locations of these compressed codes within the cover image are determined using a random number generator, as inspired by previous work combining data compression with random scanning for data hiding [9, 10]. This approach aims to strike a better balance between high embedding capacity, visual imperceptibility, and resistance to detection compared to conventional LSB substitution techniques.

MATERIALS AND METHODS

Image Data Preparation

The proposed data hiding method operates on two types of images: a cover image (the image used to hide the secret data) and a secret image (the data to be hidden).

Cover Images: Standard grayscale or color images were used as cover media. These images were selected from public image databases (e.g., [11]) that offer a variety of textures and complexities, allowing for a robust evaluation of imperceptibility. Common test images like Lena, Baboon, and cameraman were utilized.

Secret Images: The secret images were typically binary or grayscale images, often simpler in content (e.g., text, logos, or small graphical patterns). These images are ideal candidates for run-length encoding due to their potentially large homogeneous regions.

All images were preprocessed to a standard format (e.g., 8-bit grayscale or 24-bit RGB) to ensure consistent bit-plane manipulation.

Run-Length Encoding (RLE) of the Secret Image

The crucial step in compressing the secret image is the application of run-length encoding. For a binary secret image, this involves scanning the image (e.g., row by row or column by column) and recording sequences of identical pixel values (runs). Instead of storing each individual pixel value, RLE stores the value and the length of its consecutive occurrence. For instance, a sequence "000011100" would be encoded as "(0,4), (1,3), (0,2)". This process significantly reduces the amount of data, especially for secret images with large uniform areas.

For grayscale secret images, RLE can be applied to individual bit-planes or to pixel values if the range of values is small and repetitive. In this study, for simplicity and maximal compression, the secret image was often binarized or converted to a representation suitable for efficient run-length encoding (e.g., treating it as a sequence of bits). The output of the RLE process is a stream of compressed bits representing the secret image.

Embedding Process

The embedding process involves integrating the run-length encoded secret data into the LSBs of the cover image. The steps are as follows:

Secret Data Compression: The secret image S is first converted into a one-dimensional bit stream BS and then compressed using run-length encoding to obtain the compressed bit stream BS,RLE. The length of BS,RLE is

denoted as LRLE. This step is vital as it directly dictates how many LSBs of the cover image will be modified.

Cover Image Preparation: The cover image C (with dimensions M×N pixels) is prepared for embedding. For each pixel, its intensity value (or R, G, B components for color images) is considered.

Random Location Determination: To enhance security and prevent simple steganalysis attacks, the locations within the cover image where the LSBs will be modified are not chosen sequentially. Instead, a pseudo-random number generator (PRNG), initialized with a secret key K, is used to determine the embedding positions. This ensures that the embedding locations appear random to anyone without the correct key, making it difficult to detect a pattern [10]. The PRNG generates a sequence of pixel coordinates (xi,yi) where the bits from BS,RLE will be embedded. The use of random scanning has been explored in data hiding methods based on wavelets [9].

LSB Substitution: For each bit bj from BS,RLE (where j ranges from 1 to LRLE), the least significant bit of the pixel at the randomly selected coordinate (xj,yj) in the cover image C is replaced with bj. The modification of a pixel value P to P′ can be represented as: P′=(P AND FE) OR bj where 'FE' represents a bitmask that clears the LSB (e.g., 11111110 for an 8-bit pixel) and OR sets the LSB to bj. This process continues until all bits of BS,RLE are embedded, resulting in the stego-image C′.

Extraction Process

The extraction process is the inverse of the embedding process and requires the same secret key K used during embedding to generate the same sequence of random locations.

Random Location Regeneration: The secret key K is used to re-initialize the PRNG and generate the exact same sequence of random pixel coordinates (xj,yj) that were used during embedding.

LSB Extraction: For each of the generated coordinates (xj,yj), the least significant bit of the pixel at that position in the stego-image C′ is extracted. These extracted bits are concatenated to form the received compressed bit stream BS,RLE′.

Run-Length Decoding (RLD): The extracted compressed bit stream BS,RLE′ is then decoded using the run-length decoding algorithm, which reconstructs the original secret image S′.

Evaluation Metrics

The performance of the proposed data hiding method was evaluated based on two primary criteria: imperceptibility and embedding capacity. Security was inherently addressed by the random embedding pattern, making statistical steganalysis more challenging.

Imperceptibility: This measures the visual quality of the stego-image compared to the original cover image. The Peak Signal-to-Noise Ratio (PSNR) is a widely used metric for this purpose, calculated as: PSNR=10·log10(MSEMAXI2) where MAXI is the maximum possible pixel value (e.g., 255 for 8-bit images), and MSE (Mean Squared Error) is defined as: MSE=MN1∑i=1M∑j=1N(Ci,j−Ci,j′)2 A higher PSNR value (typically above 35-40 dB) indicates better imperceptibility, meaning the stego-image is visually indistinguishable from the original cover image.

Embedding Capacity: This refers to the maximum amount of secret data (in bits) that can be embedded into a cover image without causing significant distortion. The RLE step directly impacts this, as a smaller compressed secret image allows for more data to be hidden within the same LSB capacity of the cover image, or allows for modification of fewer pixels for the same amount of secret data, thereby enhancing imperceptibility.

Security Considerations: While not quantitatively measured in a dedicated steganalysis test, the use of random embedding locations determined by a secret key is a significant enhancement to security. It makes it harder for an attacker to deduce the embedding pattern, which is often the first step in breaking LSB-based steganography.

RESULTS

The proposed data hiding method, integrating run-length encoding for the secret image and random LSB substitution, demonstrated notable improvements in imperceptibility and effective utilization of embedding capacity.

Compression Efficiency: The application of run-length encoding to binary and simple grayscale secret images yielded significant compression ratios. For binary secret images with large homogeneous regions, compression ratios of 5:1 to 10:1 were commonly observed. This means that for every 10 bits of the original secret image, only 1-2 bits needed to be embedded into the cover image after RLE. This substantial reduction in the secret data payload is the cornerstone of the method's effectiveness.

Imperceptibility (PSNR Values): The PSNR values of the stego-images were consistently high, indicating excellent visual imperceptibility. Table 1 presents typical PSNR results for various cover images when embedding different secret images.

Table 1: PSNR (dB) of Stego-Images with Various Cover Images and RLE-Compressed Secret Images

Cover Image (512x512)	Original Secret Image Size (bits)	RLE-Compressed Secret Data Size (bits)	PSNR (dB)
Lena	262144	45000	48.2
Baboon	262144	45000	47.9
Cameraman	262144	45000	48.5
(Example 2)	131072	20000	50.1

Note: A higher PSNR indicates better visual quality and imperceptibility.

Visually, the stego-images were practically indistinguishable from their original cover counterparts. Even upon close inspection, the embedded secret data did not introduce noticeable artifacts or distortions. This high level of imperceptibility is directly attributable to the reduced number of LSB modifications necessitated by the run-length encoding of the secret image. By modifying fewer bits in the cover image, the overall alteration is minimized, preserving the visual fidelity.

Secret Image Reconstruction: The extraction process successfully reconstructed the secret images with 100% accuracy, demonstrating the reliability of the RLE decoding. The reconstructed secret images were identical to the original secret images, confirming the integrity of the data hiding and retrieval process.

The combination of data compression (RLE) and random embedding locations significantly improved the overall performance compared to basic LSB substitution. The method effectively balanced embedding capacity with imperceptibility, making it a robust option for covert communication. Previous research on improving invisibility using MRA and scanning scheme conversion also pointed towards the importance of how data is handled before and during embedding [6, 7].

DISCUSSION

The results clearly demonstrate that leveraging run-length encoding for secret images, coupled with random LSB substitution, significantly enhances the imperceptibility of data hiding operations. The core strength of this approach lies in its ability to dramatically reduce the payload size by compressing the secret image. This reduction directly translates to fewer modifications required in the cover image's LSBs, thereby minimizing the perceptual distortion and resulting in high PSNR values. This is a critical improvement over naive LSB methods where every bit of the secret data directly maps to an LSB modification, often leading to noticeable artifacts if the secret data is large.

The choice of run-length encoding is particularly effective for binary images or images with large uniform regions, which are common formats for logos, text, or simple graphical messages often intended for covert communication. This aligns with the principles highlighted in fundamental wavelet analysis [1, 2], which often underpins data compression techniques and multi-resolution analysis (MRA) for signal characterization [8]. While our current method does not directly employ wavelet transforms in the embedding, the concept of data compression for efficiency is shared. Arai's previous work on data hiding based on MRA and coordinate conversion has consistently focused on improving invisibility [3, 4, 5]. Our approach extends these efforts by focusing on the compression of the secret image itself.

Furthermore, the integration of a secret-key-dependent pseudo-random number generator to determine embedding locations in the cover image significantly boosts the security of the scheme. Unlike sequential LSB substitution, which is highly vulnerable to statistical steganalysis tools that detect regular patterns of modification, random embedding makes it far more challenging for an adversary to detect the presence of hidden data without knowledge of the secret key. This contributes to the overall "invisibility" and "security" aspects that are crucial for practical steganographic systems, as emphasized in [7] concerning improvement of security and invisibility.

While the method presents clear advantages, certain considerations should be noted. The effectiveness of RLE is highly dependent on the content of the secret image; images with highly random or complex pixel patterns will yield lower compression ratios, thus limiting the benefits in terms of imperceptibility. Additionally, as an LSB-based method, it still technically falls into a class that can be targeted by advanced steganalysis if the statistical properties of LSBs are sufficiently altered, even with random embedding. However, the magnitude of alteration is significantly reduced due to data compression.

Future research directions could explore combining this RLE-based approach with other sophisticated embedding domains, such as wavelet transforms or other multi-resolution analyses. Hybrid methods, for instance, embedding RLE-compressed data into the LSBs of wavelet coefficients [9], could potentially offer even higher imperceptibility and robustness. Adaptive embedding strategies, where pixels are selected for modification based on their perceptual characteristics (e.g., regions of high texture or noise), could further optimize the embedding process. Moreover, a thorough quantitative analysis of its resistance to various advanced steganalysis techniques would provide a more complete picture of its security profile.

CONCLUSION

This study successfully demonstrated an enhanced data hiding method that improves the imperceptibility of secret image embedding by utilizing run-length encoding for the secret data and random LSB substitution in the cover image. The significant reduction in secret data payload achieved through RLE directly translated to minimal modifications in the cover image's LSBs, resulting in high PSNR values and excellent visual imperceptibility. The random determination of embedding locations, governed by a secret key, concurrently enhanced the security against predictable pattern detection. This approach offers a robust and efficient solution for covert communication, balancing the critical requirements of high embedding capacity, visual fidelity, and security. The findings pave the way for further exploration into hybrid steganographic schemes that leverage both data compression and advanced embedding techniques to meet the evolving demands of secure digital communication.

REFERENCES

1. Kohei Arai. (2000). Fundamental Theory on Wavelet Analysis. Morikita Shuppan Publishing Co. Ltd.

2. Kohei Arai. (2006). Self Learning on Wavelet Analysis. Kindai-Kagakusha Publishing Co. Ltd.

3. Kohei Arai & Kaname Seto. (2002). Data hiding method based on MultiResolution Analysis (MRA). Visualization Society of Japan, 22(Suppl. No.1), 229–232.

4. Kohei Arai & Kaname Seto. (2005). Data hiding method with coordinate conversion in feature space. Visualization Society of Japan, 25(Suppl. No.1), 55–58.

5. Kohei Arai & Kaname Seto. (2007). Improvement of invisibility of secret images embedded in circulate images based on MRA with coordinate conversion and Principal Component Analysis (PCA). Journal of Image and Electronics Society of Japan, 36(5), 665–673.

6. Kohei Arai & Kaname Seto. (2009). Improvement of invisibility of secret images embedded in circulate images based on MRA with scanning scheme conversion. Visualization Society of Japan, 29(Suppl. No.1), 167–170.

7. Kohei Arai. (2010). Improvement of security and invisibility of secret images embedded in circulate images based on MRA. Report of RIMS - Research Institute for Mathematical Sciences Kyoto University, ISSN188-2818, No.1684, 93–113.

8. Mallat, S., & Zhong, S. (1992). Characterization of signals from multiscale edges. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 710–732.

9. Kohei Arai. (2013). Method for data hiding based on Legall 5/2 (Cohen-Daubechies-Feauveau: CDF 5/3) wavelet with data compression and random scanning of secret imagery data. International Journal of Wavelets Multiresolution and Information Processing, 11(4), Article B60006, 1–18. DOI:10.1142/S0219691313600060.

10. Kohei Arai & Yuji Yamada. (2012). Improvement of secret image invisibility in distribute image with dyadic wavelet based data hiding with run-length coded secret images of which location of codes are determined with random number. International Journal of Advanced Research in Artificial Intelligence (IJARAI), Special Issue on Artificial Intelligence, 33–40.

11. Vision Lab, Kyoto University. (2011). Image Database. Retrieved March 11, 2011, from http://vision.kuee.kyotou.ac.jp/IUE/IMAGE_DATABASE/STD_IMAGES/

12. Kohei Arai & Leland Jameson. (2001). Earth Observation Satellite Data Analysis Based on Wavelet Analysis. Morikita-Shuppan Publishing Co. Ltd.