Coverless Video Steganography Based on Inter Frame Combination

Liming Zou¹, Wenbo Wan¹, Bin Wei², and Jiande Sun¹(✉)

¹ School of Information Science and Engineering, Shandong Normal University, Jinan, China
² Shandong Subcenter of National Computer Network and Information Security Administrative Center, Jinan, China

Abstract. In most coverless image steganography methods, the number of images increases exponentially with the increase of hidden message bits, which is difficult to construct such a dataset. And several images in semantic irrelevance are usually needed to represent more secret message bits, which are easy to cause the attacker’s attention and bring some insecurity. To solve these two problems, a coverless video steganography method based on inter frame combination is proposed in this manuscript. In the proposed method, the hash sequence of a frame is generated by the CNNs and hash generator. To hide more information bits in one video, a special mapping rule is proposed. Through this mapping rule, some key frames in one video are selected. In the selected frames, one or several frames are used to represent a piece of information with equal length. To quickly index out the corresponding frames, a three-level index structure is proposed in this manuscript. Since the proposed coverless video steganography method does not embed one bit in video, it can effectively resist steganalysis algorithms. The experimental results and analysis show that the proposed method has a large capacity, good robustness and high security.

Keywords: Coverless information hiding · Cover selection-based information hiding · Video steganography · Multimedia security

1 Introduction

In traditional image steganography methods [1–5], secret information is embedded in images to achieve covert communication. This process of embedding information results in the image being modified. It is possible that such marked image, which is modified by traditional image steganography methods, can be detected by the state-of-the-art steganalysis algorithm under a certain payload. Based on this, the idea of coverless information hiding is proposed. “Coverless”
does not mean that no carrier is needed when secret information transmission, but represent the secret information by the carrier itself without any modification [6].

The existing coverless image steganography methods based mapping rules can be roughly divided into two categories: spatial domain and frequency domain. Spatial domain-based methods usually extract features based on pixels, and then use these feature sequences to map secret information. For example, Zou et al. proposed a method based on the average value of sub image pixels [7]. In Zou’s method, the image is first divided into several blocks and the average pixel value of each block is calculated. Then an 80-bit hash sequence is generated by comparing the average values of neighboring blocks. Finally, the 80-bit hash sequence is mapped to the secret information. This method is simple and effective, but its robustness needs to be further improved. Frequency-domain-based methods usually transform pixel domain into frequency domain firstly, and then extract corresponding features of the frequency domain. Finally, these feature sequences are mapped with secret information. For example, the DCT based method of Zhang et al. [8] and the DWT based method of Liu et al. [9]. In Zhang’s method, an 8 × 8 block discrete cosine transform of the image is performed. Then the hash sequence is generated by dc coefficients in these blocks. Finally, the dc coefficients based hash sequence is mapped to the secret information. In Liu’s method, the image is divided into 4 × 4 blocks and discrete wavelet transform is performed in each block. Then the hash sequence is generated by these DWT coefficients of low-frequency. Finally, the DWT coefficients based hash sequence is mapped to the secret information. These frequency domain-based methods greatly improve the robustness. However, these mapping based coverless image steganography methods exist two problems need to be solved. The first problem is that the image database will increase exponentially when the secret information bits hidden in an image increase. This brings great difficulty to the construction of database. The second problem is that many semantically irrelevant images are usually selected when more secret information bits need to be hidden, which brings some security challenges.

Pan et al. [10] proposed a video coverless steganography algorithm based on semantic segmentation. Specifically, the hash sequence of a video frame was generated according to the semantic segmentation network based on deep learning and the statistical histogram. The video with the key frame corresponding to the secret information was transmitted as a marked carrier. Pan et al. used video as a carrier of coverless steganography scheme firstly, and brought a new idea to the field of coverless steganography. However, these problems existing in coverless image steganography were still unsolved in the method of Pan et al.

To solve these problems mentioned above, a coverless video steganography method based on inter frame combination is proposed in this manuscript. The contributions of this manuscript are summarized as follows.

1) It is easier to construct a video dataset to the coverless information hiding task by the proposed method.
2) The proposed method greatly increases the capacity that a carrier can represent.
3) The experimental results and analysis show that the proposed method has good robustness and high security.

![Diagram of the proposed method](image)

**Fig. 1.** The framework of the proposed method.

## 2 The Proposed Method

In this section, we discuss the proposed coverless video steganography method. The framework of the proposed method is shown in Fig. 1. Firstly, we build the video dataset constructed from short videos. Secondly, hash sequences of video frames are obtained by CNNs and the hash generator. Finally, these hash sequences and video frames are built the index structure to map secret information. For the sender, the secret information is divided into multiple pieces of information with the same length. For each piece of secret information, it can be mapped one or several key frames in one video according to the mapping rule and index structure. To extract the secret information correctly, we record the ID of key frames mapped with each piece of secret information as key information and send it to the receiver along with the cover video. On the receiving end, the receiver can locate the key frames according to the key information and extract the secret message correctly.

### 2.1 Generating Hash Sequence

In this manuscript, we use the CNNs provided in [11] to extract the 4096-dimensional deep features of video frames. Then, each 4096-dimensional deep feature is blocked and fused to a feature map with length of $M$, which direction is the Zigzag-scan rule as shown in Fig. 2.
After the feature map got, we can obtain the hash sequence $H$ as shown in Eq. (1).

$$H = \{h_1, h_2, h_3, ..., h_M\}$$  \hspace{1cm} (1)

Where, $h_i (1 \leq i \leq M)$ is calculated according to Eq. (2) and Eq. (3).

$$h_i = \begin{cases} 1, & B_i > B_{i+1} \\ 0, & B_i \leq B_{i+1} \end{cases} \quad 1 \leq i \leq (M - 1) \hspace{1cm} (2)$$

$$h_i = \begin{cases} 1, & B_i > B_1 \\ 0, & B_i \leq B_1 \end{cases} \quad i = M \hspace{1cm} (3)$$

Where, $B$ represent the block of feature map. And $B_i (1 \leq i \leq M)$ follows the direction of Zigzag-scan rule.

### 2.2 Mapping Rule

To hide more information bits in one cover video, we design a mapping rule based on inter frame combination. When $M$ is defined, we know that a video frame can hide $M$-bit information. Under an ideal assumption, we hope that a video can represent $2^M$ information changes. We denote $N_1$ as the number of frames in a video. If each frame represents a $M$-bit information, then the video can represent at most $N_1$ information changes. If $N_1$ is less than $2^M$, the video can not guarantee that any length of information can be hidden. In fact, there are only a few hundred frames in some short videos, which are much less than $2^M$.

To improve the ability of a short video to hide information, we use a combination of two and three frames in the same video to represent a $M$-bit secret information. We denote $N_2$ and $N_3$ as the number of combinations of two and three frames, which are shown in Eq. (4) and Eq. (5).
The number of combinations that a video can hide $M$-bit information has expanded to $N$, as shown in Eq. (6)

$$N = N_1 + N_2 + N_3$$

We denote $H_{f_i}$ as the hash sequence of the frame with ID $= f_i$ ($1 \leq f_i \leq N_1$). The hash sequence of two and three frames is calculated by Eq. (7) and Eq. (8), where $\oplus$ is the symbol for modular two plus.

$$H_{f_i, f_j} = H_{f_i} \oplus H_{f_j}, \ 1 \leq f_i, f_j \leq N_1$$  

$$H_{f_i, f_j, f_k} = H_{f_i} \oplus H_{f_j} \oplus H_{f_k}, \ 1 \leq f_i, f_j, f_k \leq N_1$$

Based on the inter frame combination, we establish a three-level index structure to improve the retrieval efficiency. The index structure is shown in Fig. 3. It is worth mentioning that the 1st level has the highest priority and the 3rd level has the lowest priority.
2.3 Information Hiding

On the sending end, suppose there is a $L$-bit secret information needs to be hidden. The process of information hiding is illustrated as follows.

1) Dividing $L$-bit secret information into $n$ segments, which are denoted as \( \{L_1, L_2, \ldots, L_n\} \). And the length of each segment $L_t \ (1 \leq t \leq n)$ is $M$. If $L$ is not divisible by $M$, add 0 at the end. These 0 are recorded. $n$ is the number of segments as shown in Eq. (9).

\[
    n = \begin{cases} 
        \frac{L}{M}, & \text{if } L \mod M = 0 \\
        \frac{L}{M} + 1, & \text{otherwise}
    \end{cases}
\]  

(9)

2) Indexing the frames corresponding with $L_1$ according to the three-level index structure. Selecting a video with the frame corresponding with $L_1$, and recording the frame ID as the key information.

3) Based on the video selected in step 2), indexing the frames corresponding with $L_j \ (2 \leq j \leq n)$. And recording frames ID as the key information.

4) If the video selected in step 2) can not hide all $L$-bit information, return to the step 2) to select another video that can hide all $L$-bit information.

2.4 Information Extraction

On the receiving end, the receiver can extract the secret information as follow steps.

1) The receiver locates the key frames in the cover video according to the key information.

2) For each key frame combination, the deep features are extracted by CNNs. And the hash sequences are calculated by hash generator. Finally, a $M$-bit secret information is obtained according to Eq. (7) and Eq. (8).

3) Connecting all the extracted $M$-bit information in order.

4) Cropping the excess zeros to get the final $L$-bit secret information according to the zero padding record.

3 Experiment Results and Analysis

In this manuscript, we select Hollywood dataset to verify the proposed method. The Hollywood dataset contains 937 short videos with 24fps (frame per second).

3.1 Capacity

In this section, we discuss the capacity of one video frame. The capacity of one video frame depends on $M$. The comparison results with some coverless information hiding methods are shown in Table 1.
Table 1. Hidden bit number comparison in one frame.

| Methods          | [10] | [12] | [13] | [14] | Proposed method \((M = 16)\) |
|------------------|------|------|------|------|-----------------------------|
| Capacity         | 8 bit| 8 bit| 8 bit| 7–14 bit| 16 bit                     |

3.2 Robustness

In this section, we conduct the robustness experiments comparison with [10]. The attacks are shown in Table 2. In Table 2, the crf parameter is the constant rate factor of H.264 coding. The theoretical range of crf parameter is \([0, 51]\). Typically, \(crf = 18\) is considered visually lossless and \(crf = 23\) is the industry default. Therefore, we set \(crf = 18\) and \(crf = 23\) in this experiment.

Table 2. Kind of attacks used in the comparison experiment.

| Attack          | The special parameters                      |
|-----------------|---------------------------------------------|
| JPEG compression| The quality factors \(Q: 90\)                  |
| Gauss noise     | The mean \(\mu: 0\), the variances \(\sigma: 0.001, 0.005\) |
| Speckle noise   | The mean \(\mu: 0\), the variances \(\sigma: 0.01, 0.05\) |
| H.264           | Constant rate factor \(crf: 18, 23\)         |

The results of robustness experiments are shown in Table 3. The robustness is represented by the accuracy of information bit recovery.

Table 3. The accuracy of information bit recovery.

| Attack          | Size       | Literature [10] | Proposed |
|-----------------|------------|-----------------|----------|
| JPEG compression| \(Q(90)\)  | 0.8833          | 0.8589   |
| Gauss noise     | \(\sigma(0.001)\) | 0.3055          | \(0.5089\) |
|                 | \(\sigma(0.005)\) | 0.2975          | \(0.5171\) |
| Speckle noise   | \(\sigma(0.01)\) | 0.4959          | \(0.5250\) |
|                 | \(\sigma(0.05)\) | 0.3002          | \(0.5152\) |
| H.264           | \(crf(18)\) | –               | \(0.4986\) |
|                 | \(crf(23)\) | –               | \(0.4990\) |

3.3 Security Analysis

In this manuscript, no one bit is embedded in the selected carrier by the proposed method. In other words, there is no modification in the selected carrier. Therefore, the proposed method can resist all steganalysis algorithms and has high security.
4 Conclusion

A novel coverless video steganography method is proposed in this manuscript. The proposed method solves two problems that the existing coverless image steganography and coverless video steganography have not settled, which are the difficulty of constructing a dataset and the large number of transmission carriers. Besides, large capacity, good robustness and high security are achieved by the proposed coverless video steganography method.

References

1. Tian, J.: Reversible data embedding using a difference expansion. IEEE Trans. Circuits Syst. Video Technol. 13(8), 890–896 (2003)
2. Chan, C.K., Cheng, L.M.: Hiding data in images by simple LSB substitution. Pattern Recogn. 37(3), 469–474 (2004)
3. Ni, Z., Shi, Y.Q., et al.: Reversible data hiding. IEEE Trans. Circuits Syst. Video Technol. 16(3), 354–362 (2006)
4. Wang, K., Lu, Z.M., et al.: A high capacity lossless data hiding scheme for JPEG images. J. Syst. Softw. 86(7), 1965–1975 (2013)
5. Hou, D., Zhang, W., et al.: Reversible data hiding in color image with grayscale invariance. IEEE Trans. Circuits Syst. Video Technol. 29(2), 363–374 (2018)
6. Chen, X., Sun, H., Tobe, Y., Zhou, Z., Sun, X.: Coverless information hiding method based on the Chinese mathematical expression. In: Huang, Z., Sun, X., Luo, J., Wang, J. (eds.) ICICS 2015. LNCS, vol. 9483, pp. 133–143. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-27051-7_12
7. Zou, L., Sun, J., Gao, M., Wan, W., Gupta, B.B.: A novel coverless information hiding method based on the average pixel value of the sub-images. Multimed. Tools Appl. 78(7), 7965–7980 (2018). https://doi.org/10.1007/s11042-018-6444-0
8. Zhang, X., Peng, F., et al.: Robust coverless image steganography based on DCT and LDA topic classification. IEEE Trans. Multimed. 20(12), 3223–3238 (2018)
9. Liu, Q., Xiang, X., et al.: Coverless steganography based on image retrieval of DenseNet features and DWT sequence mapping. Knowl.-Based Syst. 192, 105375 (2020)
10. Pan, N., Qin, J., Tan, Y., Xiang, X., Hou, G.: A video coverless information hiding algorithm based on semantic segmentation. EURASIP J. Image Video Process. 2020(1), 1–18 (2020). https://doi.org/10.1186/s13640-020-00512-8
11. Jiang, Q.Y., Li, W.J.: Deep cross-modal hashing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3232–3240 (2017)
12. Xia, Z., Wang, X., Sun, X., Liu, Q., Xiong, N.: Steganalysis of LSB matching using differences between nonadjacent pixels. Multimed. Tools Appl. 75(4), 1947–1962 (2014). https://doi.org/10.1007/s11042-014-2381-8
13. Zheng, S., Wang, L., Ling, B., Hu, D.: Coverless information hiding based on robust image hashing. In: Huang, D.-S., Hussain, A., Han, K., Gromiha, M.M. (eds.) ICIC 2017. LNCS (LNAI), vol. 10363, pp. 536–547. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63315-2_47
14. Ruan, S., Qin, Z.: Coverless covert communication based on gif image. Commun. Technol. 50(7), 1506–1510 (2017)