An Efficient Video Steganography Scheme for Data Protection in H.265/HEVC

Hongguo Zhao[1], Menghua Pang[2], and Yunxia Liu[1](✉)

[1] College of Information Science and Technology, Zhengzhou Normal University, Zhengzhou, China
liuyunxia0110@hust.edu.cn

[2] College of Mathematics and Statistics, Zhoukou Normal University, Zhoukou, China

Abstract. An efficient and novelty video steganography scheme with adopting averting intra prediction distortion drift technique is proposed in this paper for the enhanced protection of crucial data related to H.265/HEVC digital video. With adoption of intra prediction modes selection and DCT/DST coefficients correlation, the intra prediction distortion drift can be totally prevented in intra frame (I frame). With adoption of selection of prediction patterns of multiple adjacent prediction blocks, the intra prediction distortion drift in inter frames (B frame or P frame) can be totally prevented. Compared to the previous related works, the proposed scheme applied the averting intra prediction distortion drift technique not only intra but also inter frames, and further improve the visual quality (imperceptibility) of carrier video about H.265/HEVC. Moreover, larger embedding capacity can also be achieved rather than embedding only manipulated on I frames. The experimental results have been proven the superiority of efficiency and performance about the proposed scheme.

Keywords: Video steganography · Intra distortion drift · Prediction modes · Intra/inter prediction

1 Introduction

Recently video steganography technology is becoming a powerful tool for the protection of prominent (or sensitive) data related to digital videos. Especially for the illegal distribution and propagation of movie products, video steganography can achieve a strong protection level by embedding watermark into movie products, and trace back the distribution trajectories through the network broadcasting [1]. Another supporting foundation is the continuous development of video coding technology. Since the primary video coding standards have been developed by the well-known ITU-T and ISO/IEC organizations (where H.261, H.263 is released by ITU-T, MPEG-1, MPEG-4 visual by ISO-IEC, and H.262/MPEG-2, H.264/MPEG-4 by the combination), H.265/HEVC was finalized in 2013, and now has been a key technology standard for multi-application scenarios, including scalable video coding, 3-D/stereo/multi-view video coding and the most important, the high compression efficiency for captured movies, especially for HD

D.-S. Huang et al. (Eds.): ICIC 2021, LNCS 12836, pp. 358–368, 2021.
https://doi.org/10.1007/978-3-030-84522-3_29

format products [2]. As claimed in [2, 3], for the identical perceptual video, H.265/HEVC can achieve approximately 50% bit-rate reduction compared to the proceeding standard H.264/AVC. While considering the strong needs of video products against the illegal distribution, protection of video contents [4], and the compression capability, computational resources for practical use, video steganography technology can be a desirable tool to tackle these protection problems of privacy data referred to video content or application [5].

Video steganography technology provides a potential path to protect video products from malicious spreading and legitimate rights tracing through network. Video steganography researches always searches the most optimized video signal redundancy to embed secret data into video contents (carrier videos), the changes to the carrier video introduced by embedding secret data are imperceptible for video observers except for data extractors. The existing video steganography methods can be classified into spatial and transform domain based researches according to the embedding domains. Also, based on video specific features, the video steganography researches can be classified to prediction modes [6], motion vector [7] (MV), and other specific features [8] (e.g., variable length code-VLC).

While considering the scenario of transferring video through network, transform domain based video steganography methods provides more practical application meanings. The main reason is that embedding data into spatial domains (always LSB substitution on pixels) can be easily lost after loss compression (e.g., H.264/AVC encoder). However, in the transform domain based researches, discrete cosine transform (DCT) coefficient is a basic and renowned carrier due to its majority occupying the bitstream and reversibility [4, 9–11]. Moreover, based on video specific features, intrapicture prediction modes are also a hot research field for embedding due to its fundamental role for I frame prediction and reconstruction (reference samples or decoding) process. Moreover, different from MV based methods and VLC based methods which will introduce large visual distortion for carrier videos, embedding based on prediction modes limit the distortion drift and guarantee a comparative large embedding capacity due to more textural features in I frame [7]. Based on above consideration, the combination between DCT coefficients and intrapicture prediction modes would be a potential tool to design an efficient video steganography scheme for digital video protection, especially for HD or beyond HD format videos compressed by H.265/HEVC.

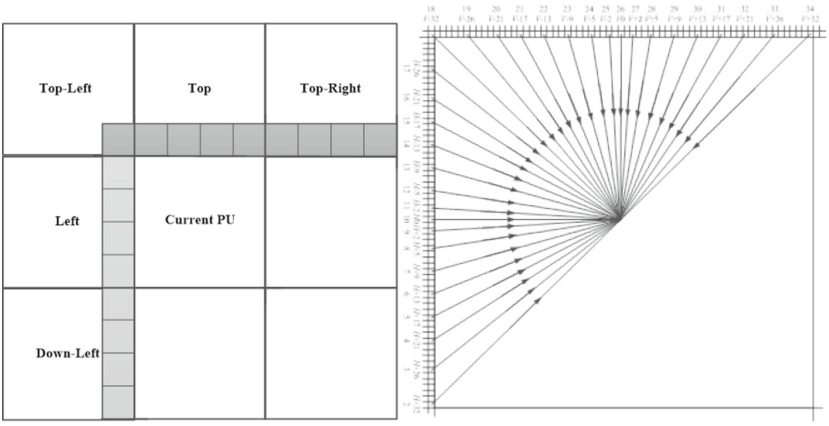
In this paper, we focus on the issues of video security protection, by designing an efficient and video steganography scheme to promote the security level of the digital video transmitted on network. Based on the relevant researches of DCT/DST (discrete sine transform) domain and intrapicture prediction modes, we propose to combine the DCT/DST coefficients and intrapicture prediction modes in 4×4 DCT/DST block to embed secret data. In order to minimize the visual distortion, we design a module to totally avert intra distortion drift (prediction modes and DCT/DST coefficient in I frame, prediction modes groups in B and P frames). Experimental evaluation has been proven that the proposed method can achieve high visual quality and security performance for carrier videos and sufficient embedding capacity for embedding.

The remainder of this paper is organized as follows: Sects. 2 reviews the related technical backgrounds of DCT/DST transformation process, prediction modes and intrapicture distortion drift. Section 3 proposes the scheme of our video steganography research based on the combination of DCT coefficients and intrapicure prediction modes in I frame and prediction modes groups in B and P frames. In Sect. 4, the experimental results are presented and evaluated our scheme. Finally, the conclusion is shown in Sect. 5.

2 Related Technical Backgrounds

The intrapicture prediction is used in I frames and B or P frames to reduce the signal spatial redundancy in H.265/HEVC. In the prediction process of I frames, the intrapicture prediction is the only prediction pattern and in B or P frames, there's probably intrapicture prediction in prediction units as the extension and supplementary for inter motion estimation. As shown in Fig. 1(a), when predicting the samples of current PU, the reference samples marked as gray will be used as reference to generate the predicted samples for current PU. Moreover, the adjacent blocks containing the reference sample is indexed in this paper as Top-Left, Top, Top-Right, Left, and Down-left blocks with their corresponding locations on current PU. The prediction angles are depicted in Fig. 2(b), where there are 33 angular prediction modes for current PU. The best prediction mode can be selected from these 33 angular prediction directions, or planar and DC modes, which is determined by the comparison of calculation of distortion (measured by SAD) and encoded bits number (entropy with CABAC). After prediction modes have been confirmed, the residual samples can be acquired by the subtraction between the original samples and the predicted samples in current PU.

The H.265/HEVC prediction modes provides higher precision than its proceeding standard H.264/AVC due to its more angular prediction directions in small prediction blocks, e.g., 4×4 prediction unit.



a) Prediction Unit and its adjacent blocks b). 33 angular prediction modes

Fig. 1. Intrapicture prediction modes and process

2.1 Intrapicture Distortion Drift in H.265/HEVC

The intrapicture distortion drift is always introduced by embedding secret data into carrier videos. Here the generation process is depicted for embedding data into DCT/DST coefficients. As shown in Fig. 1(a), if we embed secret data into the current PU's adjacent block, such as Top-Left, Top, Top-Right, Left or Down-Left. Then the reference samples, which are marked as gray region might be changed due to the embedding manipulation. However, these errors would be accumulated, and propagated to the current PU prediction process. As a result, the predicted samples would be not identical to the original predicted samples for the same current PU. The embedding errors accumulation process is defined as intrapicture distortion drift in video steganography and it would bring a considerable distortion for the carrier video.

2.2 Transformation and Inverse Transformation in H.265/HEVC

Due to the actual embedding is manipulated on DCT/DST coefficients, the main transformation and inverse transformation in H.265/HEVC is elaborated in this section. The transformation process can translate the residual samples from spatial domain to transform domain for higher precision and lower dynamic range. In H.265/HEVC, 4×4 transform block use DST transformation matrix, and other dimension blocks use DCT transformation matrix. The one-dimensional DST transformation process of 4×4 transform block can be formulated as following:

$$Y = AX \tag{1}$$

Where Y presents the transformed coefficients, X presents the residual samples after prediction process, and A indicates the transformation matrix. If the dimension of current transform block is 4×4, the transformation matrix A can be depicted as following:

$$A = \frac{2}{3}\begin{bmatrix} \sin\frac{\pi}{9} & \sin\frac{2\pi}{9} & \sin\frac{3\pi}{9} & \sin\frac{4\pi}{9} \\ \sin\frac{3\pi}{9} & \sin\frac{3\pi}{9} & 0 & -\sin\frac{3\pi}{9} \\ \sin\frac{4\pi}{9} & -\sin\frac{\pi}{9} & -\sin\frac{3\pi}{9} & \sin\frac{2\pi}{9} \\ \sin\frac{2\pi}{9} & -\sin\frac{4\pi}{9} & \sin\frac{3\pi}{9} & -\sin\frac{\pi}{9} \end{bmatrix} \tag{2}$$

Rounding and scaling above transformation matrix A, we can acquire the integer transformation matrix H and overwrite the two-dimensional DST transformation as follows:

$$H = \begin{bmatrix} 29 & 55 & 74 & 84 \\ 74 & 74 & 0 & -74 \\ 84 & -29 & -74 & 55 \\ 55 & -84 & 74 & -29 \end{bmatrix} \tag{3}$$

$$Y = HXH^T \tag{4}$$

Above Eq. (4) depicts the actual DST transformation process in H.265/HEVC. After transformation, the coefficients Y will go through post-scaling and quantization process as follows:

$$\widetilde{Y} = (Y. \times MF)/2^{(qbits+T\_Shift)} \tag{5}$$

Where $qbits = 14 + floor(QP/6)$ $MF = 2^{qbits}/Q_{step}$, and QP is the quantization parameter and Q_{step} presents the quantization step, which is determined by coding configuration and rate-distortion optimization (RDO) process for bit-rate restriction scenario.

The inverse transformation always occurs in decoding or reconstruction process. The re-scaling and inverse quantization process (Eq. 6), inverse transformation process (Eq. 7) are depicted as following:

$$Y' = \widetilde{Y} . \times Qstep. \times 2^{6-shift} \tag{6}$$

$$X' = H^T Y' H \tag{7}$$

Where $shift = 6 + floor(QP/6) - IT\_Shift$, Y' depicts the transformation coefficients after re-scaling and inverse quantization, and X' presents the acquired residual samples after inverse DST transformation process and would be used for future reconstruction samples with predicted samples.

3 Proposed Prediction Modes and DCT/DST Coefficients Based Video Steganography Scheme

The proposed video steganography scheme based on intra prediction modes and DCT/DST coefficients is illustrated in Fig. 2. The scheme can be divided into two components, including embedding and extraction process. In embedding section, appropriate 4 × 4 embedded blocks can be selected based on the intra prediction modes of the adjacent blocks corresponding to 4 × 4 current block in I frame. Meanwhile, in B and P frames, the embedded block is selected based on its adjacent blocks' prediction patterns (intra or motion estimation). Then the specific coefficients are selected for embedding according to embedding mapping rules, where different frame types meet different embedding rules. After entropy encode (CABAC or CALVC), the carrier video will be encoded to bitstream and transmitted through external network. The extraction section is an inverse loop of embedding. To guarantee the security of secret data, essential encryption and decryption are also manipulated before and after video steganography process.

3.1 Prediction Modes Selection for Embedding Blocks

The intrapicture prediction modes selection process of 4 × 4 prediction unit can be divided into two sections according different frame types (I frame, P frame and B frame). The main goal to do prediction modes selection is to totally avert intra distortion drift introduced by embedding secret data into DCT/DST coefficients. In I frames, we define three categories about adjacent blocks prediction modes, where some components need to jointly combine the specific DCT/DST coefficients to avert intra distortion drift. In B and P frames, we define one specific prediction unit where its adjacent prediction block patterns are all motion estimation. If we embedding secret data into this specific prediction unit, the distortion introduced by embedding will not accumulate to its adjacent block prediction process, which will totally avert intra distortion drift in interpicture frames.

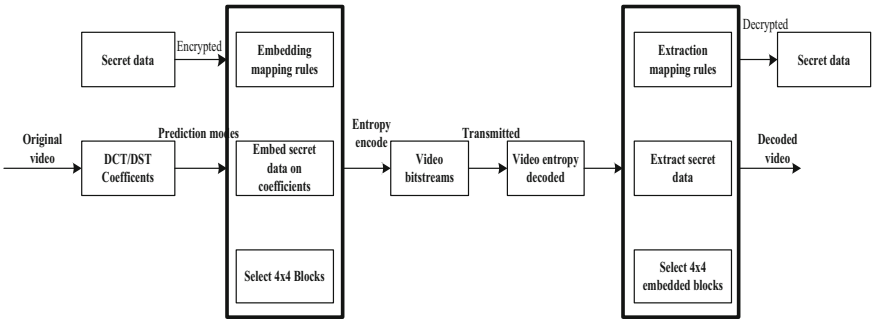


Fig. 2. Proposed video Steganography Scheme based on intra prediction modes and DCT/DST coefficients

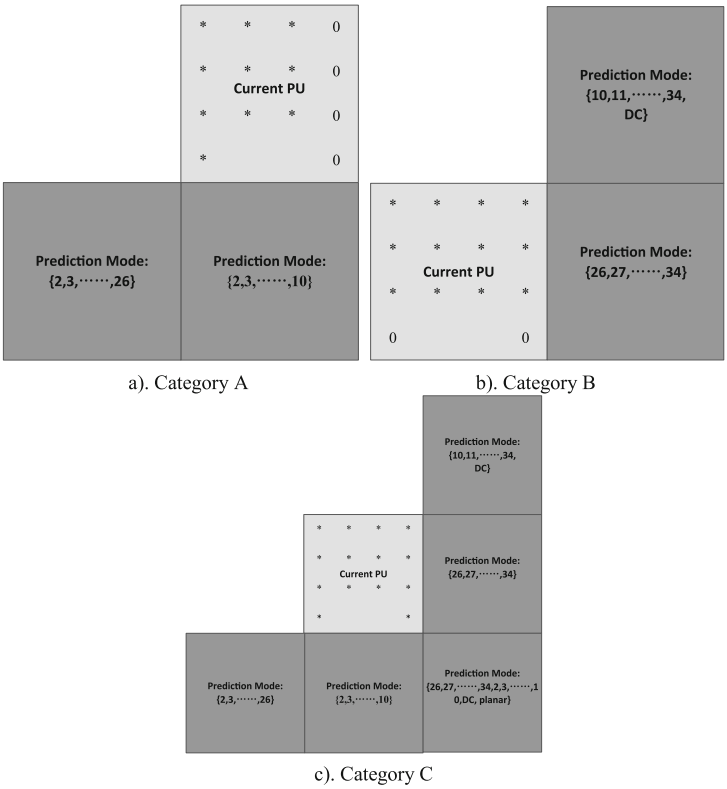


Fig. 3. Selection of prediction modes for embedding block in intrapicture frame (I frame).

As shown in Fig. 3, prediction modes can be classified into three categories based on adjacent block prediction modes in I frame. When the current PU meets category A, the embedding errors introduced by modification on current PU region will not propagated to its Down and Down-Left blocks. If we constrain the right column errors to be zero,

the embedding errors on current PU will totally prevented spreading to its adjacent prediction process. In the same way, if current PU's adjacent blocks, Right and Top-Right blocks' prediction modes meets category B, the embedding errors introduced by modification on current PU will not propagate to its Right and Top-Right blocks. If we constrain the lowest line errors to be all zero, the embedding errors generated in current PU will be totally prevented spreading to its adjacent block prediction process. In the end, if the adjacent blocks, Top-Right, Right, Down-Right, Down and Down-Left blocks prediction modes meet category C, the embedding errors introduced by embedding in current PU will not propagate to its all adjacent block prediction process.

In interpicture (B and P frames) prediction patterns, the main principle for preventing intra distortion drift is that all adjacent blocks of current block all adopts motion estimation process to generate predicted samples, then embedding secret data into current PU will not propagate to its adjacent blocks. As shown in Fig. 4, if all adjacent blocks of current blocks, Top-Right, Right, Down-Right, Down and Down-Left blocks utilize motion estimation process to generate predicted samples, the embedding errors introduced by embedding secret data into current PU will not propagate to its adjacent blocks by angle prediction directions.

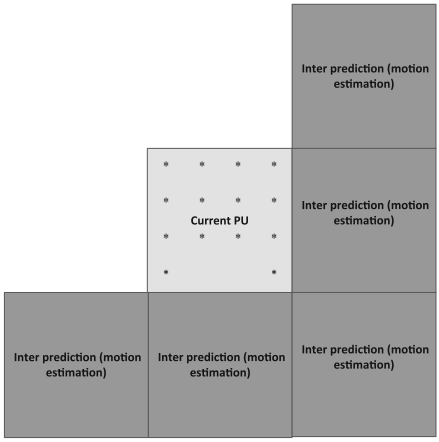


Fig. 4. Intra distortion drift when adjacent blocks all adopts motion estimation.

3.2 Specific Coefficients Selection for Embedding Blocks

As claimed in Subsect. 3.1, when current PU meets category A or B in I frames, the constrains of embedding errors to be zero should be imposed to avert intra distortion drift, which can be achieved by the selection operation on specific coefficients on current embedding block. The operation of embedding data is done on quantized DCT/DST coefficients, so the embedding error E can be presented according to formulas (6) and (7) as follows:

$$E = H^T(\Delta. \times Qstep. \times 2^{6-shift})H \tag{8}$$

Where Δ presents the actual modification on quantized DCT/DST coefficients. When current PU meet category A, the rightest column intensity values of embedding error E should be enforced to all zero, which can be represented as

$$E = \begin{bmatrix} e_{00} & e_{01} & e_{02} & 0 \\ e_{10} & e_{11} & e_{12} & 0 \\ e_{20} & e_{21} & e_{22} & 0 \\ e_{30} & e_{31} & e_{32} & 0 \end{bmatrix} \tag{9}$$

Combining Eq. (8) and (9), we can find the modification on DCT/DST coefficients should meet the following principle:

$$\Delta = \begin{bmatrix} \delta_{00} & 0 & -\delta_{00} & \delta_{00} \\ \delta_{10} & 0 & -\delta_{10} & \delta_{10} \\ \delta_{20} & 0 & -\delta_{20} & \delta_{20} \\ \delta_{30} & 0 & -\delta_{30} & \delta_{30} \end{bmatrix}$$

In which, if we embed data '1' into δ_{00},we should correspondingly embed data '-1' and '1' into coefficients δ_{02} and δ_{03}.

Similarly, When current PU meet category B, the lowest row intensity values of embedding error E should be enforced to all zero, which can be represented as

$$E = \begin{bmatrix} e_{00} & e_{01} & e_{02} & e_{03} \\ e_{10} & e_{11} & e_{12} & e_{13} \\ e_{20} & e_{21} & e_{22} & e_{23} \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{10}$$

Combining Eq. (8) and (10), we can find the modification on DCT/DST coefficients should meet the following principle:

$$\Delta = \begin{bmatrix} \delta_{00} & \delta_{01} & \delta_{02} & \delta_{03} \\ 0 & 0 & 0 & 0 \\ -\delta_{00} & -\delta_{01} & -\delta_{02} & -\delta_{03} \\ \delta_{00} & \delta_{01} & \delta_{02} & \delta_{03} \end{bmatrix}$$

In which, if we embed data '1' into δ_{00},we should correspondingly embed data '-1' and '1' into coefficients δ_{20} and δ_{30}.

Extraction module is an inverse process compared to embedding module. During the decoding process, in I frame, the current PU's quantized DCT/DST coefficients and its adjacent blocks prediction modes should be obtained from the selection category types with the same constraints as embedding. However, in interpicture frames (B and P frames), we only need to know the current PU and its adjacent blocks prediction patterns (intra prediction modes or motion estimation). Based on above pre-works, we can exactly extract the embedded secret data from corresponding DCT/DST coefficients with the same operation as embedding module.

4 Experimental Evaluation

The proposed video steganography scheme is manipulated and evaluated with the H.265/HEVC reference software HM16.0, including several public test video samples from resolution in the range of 416 × 240 to 1920 × 1080. The coding parameters are set as follows: frame-rate is set to be 30 frames/s, quantization parameter is set to be 32, and the test video sequence is set to be all intra frames, B and P frames with the interval 4. The main evaluation includes PSNR, embedding capacity and bit-rate increase, and the actual experiment results of this work is represented as following:

The subjective visual quality about our proposed scheme is depicted in Fig. 5, where the first column are the original video samples resoluations in the range of 416 × 240 to 1280 × 720 (BasketballPass: 416 × 240, RaceHorses: 832 × 480 and KristenAndSara: 1280 × 720), the second column are the compressed video samples without video steganography and the third column are the proposed video steganography scheme. It can be seen that our proposed scheme has achieved good visual quality on carrier videos compared to the reference samples without embedding. Figure 6 provides comparisons about PSNR and bitrate when we embed various volumes of secret data in test video sample BasketballPass. The experimental results have proved a good embedding performance of our proposed scheme on PSNR and bit-rate increase.

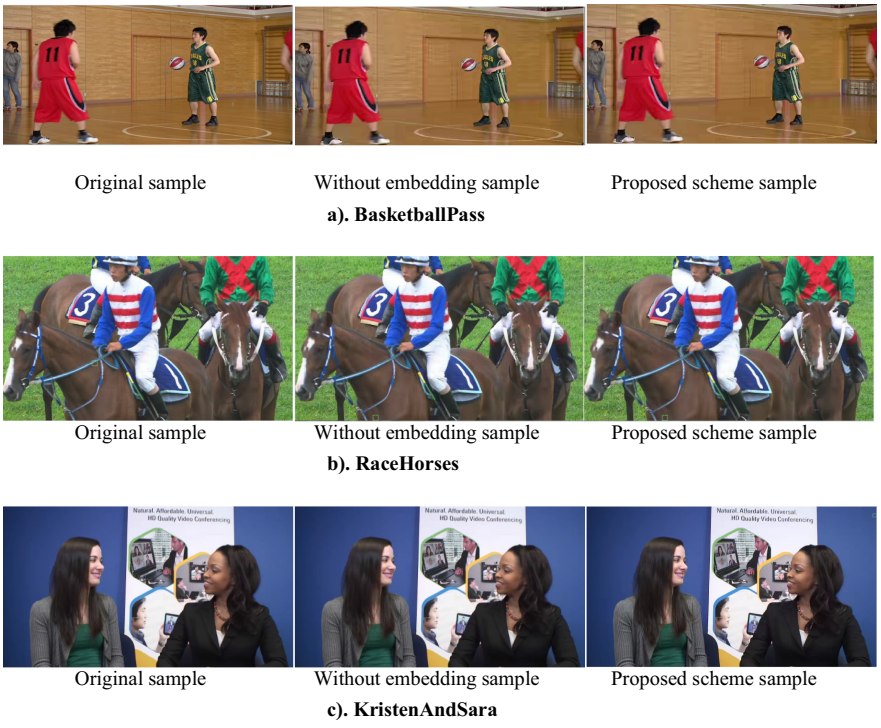


Fig. 5. Subjective visual quality of the proposed video steganography scheme

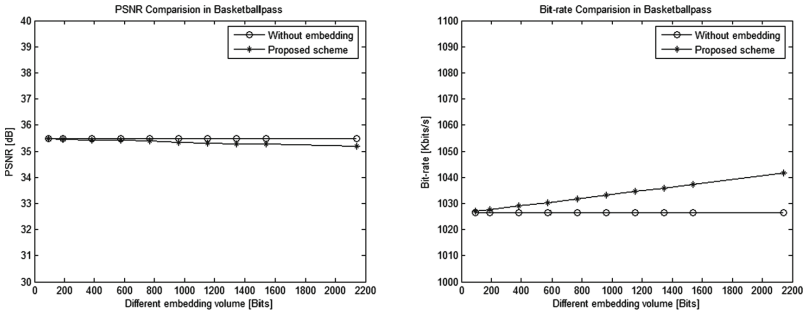


Fig. 6. Comparisons of PSNR and bit-rate with embedding different volume secret data.

Table 1. Performance of the proposed method

Video sequence	PSNR	Proposed method		
		PSNR	Capacity (bits)	Bit-rate increase
BasketballPass	35.48	35.36	2144	0.63%
RaceHorses	34.7	34.4	10092	1.20%
KristenAndSara	46.38	45.11	10028	1.41%

Table 1 provides the embedding performance of our proposed video steganography method, where 20 frames are used to test the PSNR, embedding capacity and bit-rate increase. For the visual quality evaluation, PSNR is the average value of all tested frames, and bit-rate increase is the comparison between the compressed video bitsteams with and without embedding secret data. It can be seen that for visual quality, the PSNR of our proposed scheme is 35.36 dB, 34.4 dB, and 45.11 dB for tested videos BasketballPass, RaceHorses and KristenAndSara. However, the original PSNR values are 35.48 dB, 34.7 dB and 46.38 dB, respectively. For embedding capacity, our proposed scheme has achieved 2144 bits, 10092 bits, and 10028 bits, respectively. For bit-rate increase, our proposed scheme is 0.63%, 1.2% and 1.41%, respectively. It can be seen from Table 1 that our proposed scheme can achieve a good performance on visual quality, embedding capacity and bit-rate increase.

5 Conclusion

In this paper, an effective video steganography scheme based on selection of intrapicture prediction modes is proposed for video secret data protection. The proposed scheme mainly utilize the three categories of prediction modes in I frames, prediction patterns in B and P frames, and corresponding DCT/DST coefficients to embed secret data. The experimental results show that our proposed scheme can achieve a good embedding performance on visual quality, embedding capacity and bit-rate increase. The proposed video steganography can provide a strong support for the protection of copyright of

digital videos, an effective trace back tool for malicious distribution of video products based on H.25/HEVC coding standard.

Acknowledgment. This paper is sponsored by the National Natural Science Foundation of China (NSFC, Grant No. 61572447).

References

1. Asikuzzaman, M., Alam, M.J., Lambert, A.J., Pickering, M.R.: Imperceptible and robust blind video watermarking using chrominance embedding: a set of approaches in the DT CWT domain. IEEE Trans. Inf. Forensics Secur. **9**(9), 1502–1517 (2014)
2. Sullivan, G.J., Ohm, J.R., Han, W.J., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. IEEE Trans. Circuits Syst. Video Technol. **22**(12), 1649–1668 (2012)
3. Ohm, J.R., Sullivan, G.J., Schwarz, H., Tan, T.K., et al.: Comparison of the coding efficiency of video coding standards-Including high efficiency video coding (HEVC). IEEE Trans. Circuits Syst. Video Technol. **22**(12), 1669–1684 (2012)
4. Liu, Y.X., Zhao, H.G., Liu, S.Y., et al.: A robust and improved visual quality data hiding method for HEVC. IEEE Access. **6**, 53984–53987 (2018)
5. Zhang, X.P.: Reversible data hiding with optimal value transfer. IEEE Trans. Multimedia. **15**(2), 316–325 (2012)
6. Liu, Y., Jia, S., Hu, M., et al.: A reversible data hiding method for H.264 with Shamir's (t, n)-threshold secret sharing. Neurocomputing **188**, 63–70 (2016)
7. Yang, J., Li, S.: An efficient information hiding method based on motion vector space encoding for HEVC. Multimedia Tools Appl. **77**(10), 11979–12001 (2017). https://doi.org/10.1007/s11042-017-4844-1
8. Zhao, H.G., Liu, Y.X., Wang, Y.H., et al.: A video steganography method based on transform block decision for H.265/HEVC. IEEE Access. **9**, 55506–55521 (2021). https://doi.org/10.1109/ACCESS.2021.3059654
9. Yao, Y.Z., Zhang, W.M., Yu, N.H.: Inter-frame distortion drift analysis for reversible data hiding in encrypted H.264/AVC video bitstreams. Signal Process. **128**, 531–545 (2016)
10. Zhao, H., Pang, M., Liu, Y.: Intra-frame adaptive transform size for video steganography in H.265/HEVC bitstreams. In: Huang, D.-S., Premaratne, P. (eds.) Intelligent Computing Methodologies: 16th International Conference, ICIC 2020, Bari, Italy, 2–5 October 2020, Proceedings, Part III, pp. 601–610. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-60796-8_52
11. Zhao, H.G., Liu, Y.X., Wang, Y.H., Wang, X.M., Li, J.X.: A blockchain-based data hiding method for data protection in digital video. In: International Conference on Smart Blockchain: SmartBlock 2018, Tokyo, Japan, pp. 99–110 (2018)