VIDEO STEGANALYSIS BASED ON THE CONSTRAINTS OF MOTION VECTORS

Xikai Xu, Jing Dong, Wei Wang and Tieniu Tan

Center for Research on Intelligent Perception and Computing,
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
E-mail: xikaixu@gmail.com, { jdong, wwang, tnt }@nlpr.ia.ac.cn

ABSTRACT

In this paper, we focus on detecting data hiding in motion vectors of compressed video and propose a new steganalytic algorithm based on the mutual constraints of motion vectors. The constraints of motion vectors from multiple frames are analyzed and formulized by three functions, then statistical features are extracted based on these functions. Moreover, we also incorporate calibration method to improve the detection accuracy. Experimental results demonstrate that the proposed method can effectively attack typical motion-vector-based video steganography.

Index Terms— Video steganalysis, stegnography, data hiding, motion vector, mutual constraints

1. INTRODUCTION

Data hiding in video stream can achieve relatively high payload of secret messages without drawing any suspicions. Since video resources on the Internet are becoming abundant nowadays, many steganographic algorithms for video have been proposed. As videos are often stored and transmitted in compressed format, the hiding messages should survive video lossy compression. Therefore, most data hiding schemes carry out in the compressed domain by modifying DCT coefficients [1, 2, 3, 4] or motion vectors (MVs) [5, 6, 7, 8, 9, 10] or by altering some details of video encoding [11, 12, 13]. Because of the advantage of much less distortion to the visual quality of the reconstructed frames, MV-based video steganography become popular recently. In this paper, we focus on detecting this kind of data hiding in videos.

MVs are generated in the process of motion estimation while compressing a video. Most steganographic algorithms hide data in MVs by changing their attributes such as their magnitudes, phase angles, etc. And the key issue is the selection of candidate MVs for data hiding under the principle of maintaining the robustness and low distortions. In [6, 7], the MV whose magnitude is above a predefined threshold is selected and its least signicant bits (LSB) of both components are used for hiding the secret message bit stream. The authors in [8] embed each bit in a pair of MVs by utilizing the difference of their phase angles. The method in [9] focused on achieving a minimum distortion to the prediction error by choosing MVs associated with macroblocks of high prediction error. In [14], the authors use adopting nonshared rules to select MVs and minimize the distortion by using perturbed motion estimation.

Correspondingly, video steganalysis is drawing more attention while video steganographic algorithms are constantly emerging. Similar with image steganalysis, the basic assumption for video steganalysis is that the embedding of a message changes some statistical properties of the video. Hence the essence of steganalytic methods is to reveal and measure these distortions. For this task, pattern classication is employed, in which discriminative features are extracted from cover and stego videos and a classier as detector is trained using machine learning methods. Inspired by image steganalysis, many feature sets are built for detecting data hiding in video frames, but these methods cannot effectively attack MV-based steganography since the modifications of MVs produce little quality degradation to the frames. Thus, it is desirable to design specific methods to detect such MV-based data hiding scheme. In [15], on the distribution of the differences between adjacent MVs, the authors formulated feature vector by aliasing degrees of the probability mass function (PMF) and the center of mass (COM) of the characteristic function. Later, this method was improved in [16] by using second-order difference of adjacent MVs. But these features were incompetent to detect new hiding algorithms which employ restoration strategies to maintain the statistics such as the method in [10]. The authors in [17], pointed out the phenomenon of MV reversion and proposed calibration based features for steganalysis. However, their features are sensitive to the tendency of MV reversion and the detection performance is likely to drop if some optimized measures were adopted to weaken the inclination of MV reversion.

In this paper, we propose a new method to detect the existence of hidden messages in MVs based on the mutual constraints of MVs from multiple frames, provided the fact that the constraints will be destroyed in the process of data hiding and it is very difficult for the data hiding algorithms to

maintain all these constraints. Experimental results show that our method outperforms the state-of-the-art steganalyzer on MV-based steganography.

The remainder of this paper is organized as follows: in Section 2 we overview the process of video compression and then discuss the mutual constraints of MVs. In Section 3, we measure the deviation of the constraint of MVs introduced by data hiding and build features for steganalysis. Then the experimental results are presented in Section 4. Finally, the conclusion is drawn in Section 5.

2. THE CONSTRAINTS OF MOTION VECTORS

In this section, we introduce briefly the basic principle and process of video compression first, and then analyze the details of motion-compensated prediction to explain the constraints of motion vectors.

2.1. Video compression

Video compression is used to reduce spatial and temporal redundancy in video data. Many international standards are set for video compression, such as MPEG-2, MPEG-4, ITU-T H.263, H.264, etc. and a variety of video compression codecs are designed based on these standards. Although the codecs are varying in details, they are similar in using two basic techniques: intraframe coding and interframe coding [18]. The intra-coded frame is known as I-frame, which is encoded using image compression techniques similar to JPEG. Meanwhile, interframe coding typically uses motion estimation and compensation to predict current frame by one or more neighboring frames. The frame predicted from previous frames is called P-frame, while the frame predicted bidirectionally from previous and future frames is called B-frame. Usually, the prediction is block-based: frames are divided into non-overlapping macroblocks. Then the encoder will try to find a block similar to the one it is encoding on a previously encoded frame, referred to as a reference frame. In most designs, this process is done by a block matching algorithm. If the encoder succeeds in finding a matching block on a reference frame, it will obtain prediction errors (residuals) and a motion vector (with horizontal and vertical components) pointing to the matched block. Fig.1. shows the whole process.

When compressing a video, the majority of video codecs arrange I-frames, B-frames and P-frames into GOPs (Group Of Pictures). Each GOP begins with an I-frame and then comes a P-frame after every few B-frames. In MPEG-2 standard, a typical GOP structure is IBBPBBPBBI. The I-frame is used to predict the first P-frame and these two frames are also used to predict the first and the second B-frame. The second P-frame is predicted using the first P-frame and they join to predict the third and fourth B-frames. The fifth and sixth B-frames are predicted using the second P-frame and the second I-frame. The scheme is shown in Fig.2.

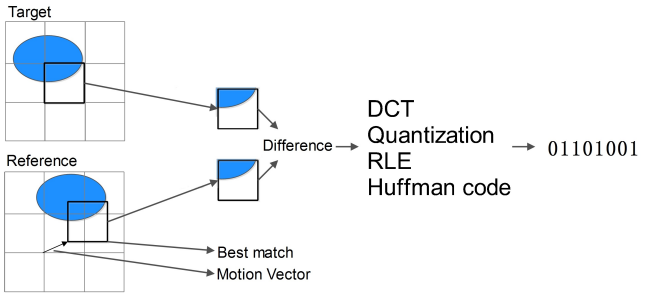

Fig. 1. Interframe coding process.

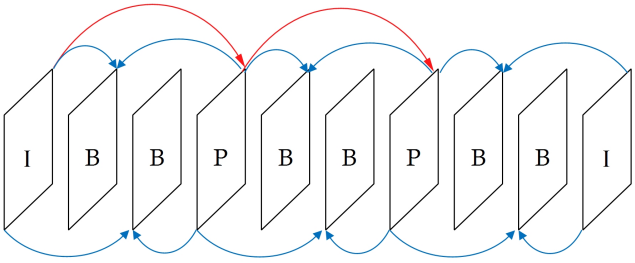

Fig. 2. Frames in a GOP and their relationships.

2.2. Constraints of MVs

To illustrate the constraints of MVs from different frames, we focus on the frames in a GOP with a certain structure IBBPBBPBBI as shown in Fig.2. For example, given a video sequence of a moving man, and the first four frames in a GOP are shown in Fig.3. In the motion compensation stage of encoding, P4 is predicted by I1 while B2 and B3 are bidirectional predicted using I1 and P4. As the prediction is block-based, we consider a certain block (for instance, the block contains the man's head) in current being encoded frame, as shown in Fig.4. Assume that MB_{I1}, MB_{P4}, $MB_{B1}$and MB_{B2} are similar blocks in different frames: I1, P4, B1 and B2, respectively. When P4 is being encoded, for the block MB_{P4} in P4, the encoder will find the best matching block MB_{I1} in I1 and return a motion vector $\mathbf{mv_0}$, as shown in Fig4.(a). Similarly, for block MB_{B1} in B1 and MB_{B2} in B2, the matching block in I1 and P4 are MB_{I1} and MB_{P4}, the corresponding forward and backward MVs are $\mathbf{mv_1}$ and $\mathbf{mv_2}$, $\mathbf{mv_3}$ and $\mathbf{mv_4}$, as shown in Fig4.(b) and (c).

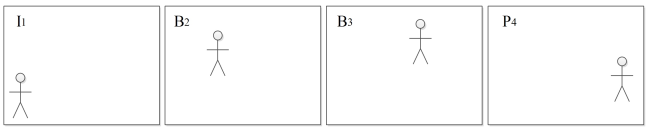

Fig. 3. The first four frames in a GOP of a moving man.

We consider these MVs in the same coordinate system, as shown in Fig.5. Then we can find the constraints of these

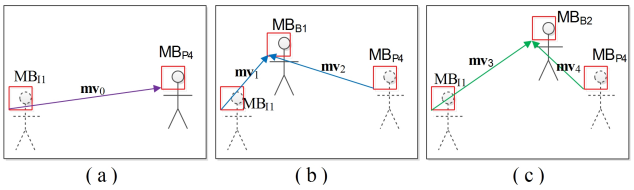

Fig. 4. Block-based motion estimation. (a) Matching block for MB_{P4} (b) Matching blocks for MB_{B1} (c) Matching blocks for MB_{B2}.

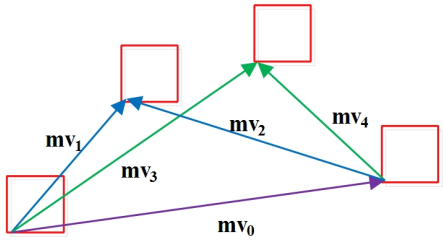

Fig. 5. The constraints of four MVs.

motion vectors easily. According to the operation rule of vector (the Triangle Rule), the following equations can be established

$$\mathbf{mv}_2 = \mathbf{mv}_1 - \mathbf{mv_0} \tag{1}$$

$$\mathbf{mv}_4 = \mathbf{mv}_3 - \mathbf{mv_0} \tag{2}$$

From (1) and (2), we can also get

$$\mathbf{mv_1} - \mathbf{mv_2} = \mathbf{mv_3} - \mathbf{mv_4} \tag{3}$$

In order to state conveniently, we take 4 frames, two anchor frames (I or P) and two B frames between them, as a motion estimation group (MEG), and there are three MEGs in a GOP. As frames are divided into non-overlapping blocks in motion compensation, we also take four matched blocks from each frame in a MEG, as a group of blocks (GOB). Five MVs can be obtained in the matching process of the blocks in a GOB. Then equations (1), (2) and (3) indicate the constraints of the MVs of a GOB.

However we should note that these equations can be true only when the four blocks in a GOB are similar blocks and the encoder succeeds in matching them.

3. PROPOSED METHOD

3.1. MV constraint-based features

As we revealed in Section 2, the MVs of a GOB have constraints. It is expected that the constraints will be destroyed by the data hiding operation. Based on this assumption, discriminative features can be extracted from cover and stego videos for training a classier to detect data hiding in MVs.

We calculate the difference (also is a vector, as follows) between the two sides of each constraint equation, and use the L1 norm of the difference vector to measure the deviation of the constraint introduced by data hiding operation.

$$d_1 = \|\mathbf{mv_0} + \mathbf{mv}_2 - \mathbf{mv}_1\|_1 \tag{4}$$

$$d_2 = \|\mathbf{mv_0} + \mathbf{mv}_4 - \mathbf{mv}_3\|_1 \tag{5}$$

$$d_3 = \|\mathbf{mv}_1 + \mathbf{mv}_4 - \mathbf{mv}_2 - \mathbf{mv}_3\|_1 \tag{6}$$

Taking each GOP as a unit of analysis, we search all GOBs in a GOP and then calculate d_1, d_2 and d_3 for each GOB. We make use of the histograms (defined in the following) of d_1, d_2 and d_3 in a GOP as steganalytic features.

$$H_1(k) = |\{i\,|d_1(i) = k\}| \tag{7}$$

$$H_2(k) = |\{i\,|d_2(i) = k\}| \tag{8}$$

$$H_3(k) = |\{i\,|d_3(i) = k\}| \tag{9}$$

Where $k = 0,\ 1, \cdots,\ U,\ \ i = 1,\ 2, \cdots,\ N$, U is the upper bound of the deviation and N is the number of GOBs in one GOP. We predefine a threshold T and only use histograms that $k \leq T$, because histograms of $k{>}T$ are usually statistically unimportant due to the small number of them. Since H_1 and H_2 are very similar, we adopt the average of them to reduce the feature dimension, and then we combine it with H_3 to get a merged feature vector as

$$F = [(H_1 + H_2)/2 \quad H_3] \tag{10}$$

3.2. Calibration of MVs

Calibration is used to estimate statistical features of the cover from the stego media. It has been shown to improve detection accuracy of feature-based blind steganalysis. In [17], the authors explained the reversion of MVs while recompressing the stego video and then suggested a calibration-based approach for video steganalysis. Stego video is perceptually similar to the cover video. While recompression, the encoder recalculates the MVs and the obtained MVs are hardly influenced by previous modification (data hiding). The MVs of the recompressed stego video approximate to that of the original cover video.

In our method, we use the calibrated MVs to calculate the deviations $d_1^{'}$, $d_2^{'}$, and $d_3^{'}$ as described in (4, 5, 6). The histograms of the differences between $d_j^{'}$ and $d_j (j = 1, 2, 3)$ are used as features and they will be more sensitive to the changes of MVs introduced by data hiding. So we combine these features with the initial feature vector F to obtain a new feature vector for steganalysis.

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed method, we compare it with the state-of-the-art method (Cao *et al.*'s

Table 1. Experimental results

	Steg1			Steg2		
	AR(%)	TN(%)	TP(%)	AR(%)	TN(%)	TP(%)
Cao's method	86.14	87.31	84.97	72.36	71.47	73.25
Proposed method	89.29	90.33	88.24	80.21	81.32	79.10

method) reported in [17] on attacking two typical MV-based steganographic algorithms: Xu *et al*.'s [7] and Aly's [9], referred to as Steg1 and Steg2. The video database is composed of 28 CIF video sequences in the 4:2:0 YUV format downloaded from the Internet. The cover and stego videos are all compressed using MPEG-2 encoder with standard settings (the GOP structure is IBBPBBPBBI). Feature vector is extracted from each GOP and the total number of GOPs is up to 1350. We use SVM with RBF kernel as classifier and randomly select 20 video sequences (931 GOPs) as training set and the rest as testing set.

The experimental results are shown in Table 1. We use true positive (TP) rate, true negative (TN) rate, and average rate (AR) to compare the detection performance. True positive rate stands for proportion of stego samples be correctly classified, and vice versa true negative rate. Average rate is the average value of TP and TN. Note that all of the rates are calculate treating each GOP as a unit. The detector receiver operating characteristic (ROC) curves of the detectors are plotted in Fig.6.

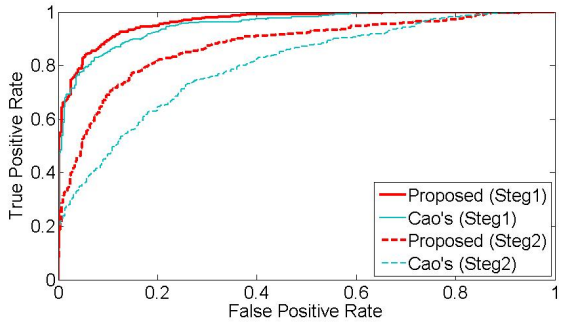

Fig. 6. ROC curves of the detectors.

From the experimental results, it is clear that our proposed method outperforms Cao's especially for Steg2. As mentioned in Section 2, Steg1 chooses the MVs with large magnitudes for data hiding and it is relatively easy to be detected; Steg2 is a new method which has very low distortion to the quality of the video and it is believed as one of the most undetectable MV-based steganographic algorithms. Our method shows obvious advantage in attacking Steg2.

5. CONCLUSION

This paper has introduced a new method against MV-based video steganography. The method has shown higher performance compared to another one from the literature. Although our method needs to be tested in attacking more steganographyic algorithms, it is convinced that the constraints of MVs are helpful for steganalysis. This new idea might suggest new issues for working in this field. It is expected that the constraints of MVs may play a greater role in the future if we can find a better way to describe and measure the statistical changes of such constraints.

6. ACKNOWLEDGEMENTS

This work is supported by the National Basic Research Program of China (Grant No. 2012CB316300) and the National Key Technology R&D Program (Grant No. 2012BAH04F02).

7. REFERENCES

[1] G. Caccia and R. Lancini, "Data hiding in mpeg-2 bit stream domain," in *Proc. Int. Conf. Trends in Communications*. IEEE, 2001, vol. 2, pp. 363–364.

[2] Y. Wang and E. Izquierdo, "High-capacity data hiding in mpeg-2 compressed video," in *Proc. of the 9th International Workshop on Systems, Signals and Image Processing*, 2002, pp. 212–218.

[3] D. Stanescu, M. Stratulat, B. Ciubotaru, D. Chiciudean, R. Cioarga, and M. Micea, "Embedding data in video stream using steganography," in *Proc. Int. Symp. Applied Computational Intelligence and Informatics*. IEEE, 2007, vol. 1, pp. 241–244.

[4] B. Wang and J. Feng, "A chaos-based steganography algorithm for h. 264 standard video sequences," in *Proc. Int. Conf. Communications, Circuits and Systems*. IEEE, 2008, pp. 750–753.

[5] F. Jordan, M. Kutter, and T. Ebrahimi, "Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video," *ISO/IEC Doc. JTC1/SC29/WG11 MPEG97/ M*, vol. 2281, pp. 27–31, 1997.

[6] Cong Van Nguyen, D.B.H. Tay, and Guang Deng, "A fast watermarking system for h.264/avc video," in *Proc. Asia Pacific Conference on Circuits and Systems*. IEEE, 2006, pp. 81–84.

[7] C. Xu, X. Ping, and T. Zhang, "Steganography in compressed video stream," in *Proc. Int. Conf. Innovative Computing, Information and Control*. IEEE, 2006, vol. 1, pp. 269–272.

[8] Ding Yufang and Long Wenchang, "Data hiding for digital video with phase of motion vector," in *Proc. Int. Symp. Circuits and Systems*. IEEE, 2006.

[9] H.A. Aly, "Data hiding in motion vectors of compressed video based on their associated prediction error," *IEEE Transactions on Information Forensics and Security*, vol. 6, pp. 14–18, Mar. 2011.

[10] Huiyun Jing, Xin He, Qi Han, and Xiamu Niu, "Motion vector based information hiding algorithm for h.264/avc against motion vector steganalysis," in *Proc. Asian conference on Intelligent Information and Database Systems*. 2012, vol. 8, pp. 91–98, Springer-Verlag.

[11] S.K. Kapotas, E.E. Varsaki, and A.N. Skodras, "Data hiding in h. 264 encoded video sequences," in *Proc. IEEE 9th Workshop on Multimedia Signal Processing*. IEEE, 2007, pp. 373–376.

[12] Gaobo Yang, Junjie Li, Yingliang He, and Zhiwei Kang, "An information hiding algorithm based on intra-prediction modes and matrix coding for h.264/avc video stream," *AEU - International Journal of Electronics and Communications*, vol. 65, no. 4, pp. 331 – 337, 2011.

[13] T. Shanableh, "Data hiding in mpeg video files using multivariate regression and flexible macroblock ordering," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 455–464, april 2012.

[14] Y. Cao, X. Zhao, D. Feng, and R. Sheng, "Video steganography with perturbed motion estimation," in *Proc. Information hiding*, 2011, vol. 6958, pp. 193–207.

[15] Yuting Su, Chengqian Zhang, and Chuntian Zhang, "A video steganalytic algorithm against motion-vector-based steganography," *Signal Process*, vol. 91, pp. 1901–1909, Aug. 2011.

[16] Yu Deng, Yunjie Wu, Haibin Duan, and Linna Zhou, "Digital video steganalysis based on motion vector statistical characteristics," *Optik*, 2012.

[17] Yun Cao, Xianfeng Zhao, and Dengguo Feng, "Video steganalysis exploiting motion vector reversion-based features," *Signal Processing Letters*, vol. 19, no. 1, pp. 35–38, Jan. 2012.

[18] T. Sikora, "Mpeg digital video-coding standards," *Signal Processing Magazine*, vol. 14, pp. 82 –100, Sept. 1997.