Scientific Journal Impact Factor (SJIF): 1.711

e-ISSN: 2349-9745
p-ISSN: 2393-8161

International Journal of Modern Trends in Engineering and Research
www.ijmter.com

Design and Implementation of Data Hiding Technique by Using MPEG Video with Constant Bit Rate

Ramakrishna Hegde[1], Dr.Jagadeesha S[2]
[1]*Department of Computer Science and Engineering, SDM Institute Of Technology, Ujire*
[2]*Department of Electronics and Communications Engineering, SDM Institute of Technology, Ujire*

Abstract-This paper proposes a technique on data hiding approaches using compressed MPEG video files. This approach hides the message bits by modulating the quantization scale of constant bit rate MPEG videos. Payload is calculated for each macroblock and proposes to achieve one message bit per macroblock. Macroblock level feature variables are calculated.To find the association between macroblock level feature variables and value of a hidden message bit, a Second Order Multivariate regression model is used. To achieve the very high prediction accuracy, the regression model is used by the decoder. To decode the message, a feature variable of MBs from the encoded bit stream are computed by the decoder and expands them to the second order and uses the model weights to predict the message bits. This solution provides very high precision accuracy in predicting the message bits . The proposed technique is analyzed in term of quality distortion, excessive bit rate, message pay load and message extraction accuracy. The proposed solution is better in terms of message payload while causing the less distortion and reduced compression overheads compare to the previous works.

Keywords : Steganography, Staganalysis, MPEG video, Macroblocks, Quantization scale modulation.

I INTRODUCTION

In many new applications for military and civilian purpose, the contributions of steganography are immense. As people become aware of the internet day-by-day, the number of users in the network increases considerably thereby, facing more challenges in terms of data storage and transmission over the internet, for example information like account number, password etc. Hence, in order to provide a better security mechanism, we are introducing efficient data hiding techniques called steganography. Steganography comes from the Greek word meaning covered writing. Steganography can be defined as the hiding of the message within another so that presence of hidden message is indiscernible. The key concept behind steganography is that the message to be transmitted is not detectable to the casual eye. In fact, who are not intended to be the recipients of the message should not even suspect that a hidden message exists.

Steganography is an efficient technique to provide secure data transmission over a communication channel. In the view of security, steganography does not allow to detect the presence of secrete data present in a carrier file. Carrier files may be of any type like video(any format), text, audio files, still images etc.,. In data hiding technique the secrete message is hiding into a compressed video bit steam for copy right protection, access control, content annotation and transaction tracking. The authors of [1] used data hiding to enable real time scene change detection in compressed video. For instance, [2] used data hiding techniques to assess the quality of compressed video in the absence of the original reference.

In modern digital steganography, data is first encrypted by the usual means and then inserted using the special algorithm, into redundant data that is part of a particular file format. By applying the encrypted data that this redundant data in some random or nonconspicuous way, the result will be data that appears to have the “noise” patterns of regular, non encrypted data.

In general, for data hiding, existing solutions rely on hiding message bits in Descrete Cosine Transformation (DCT) coefficients [3[4], motion vectors(MVs)[5][6][7], quantization scale[8] or prediction

modes. Data hiding can also be applied before compression. For example [3] introduced a method that is robust to heavy JPEG compression. It is also possible to hide data in wavelet domain as reported in [4].There is an much improved data hiding technique based on BCH(n,k,t) coding which is reported in [9].

Recently as reported in [10]MPEG video files can be used to hide the data using Multivariate regression and flexible macroblock ordering. Wang and Moulin [11] have shown that, risk of detection of data can be reduced to zero level with effective steganography, as long as embedder has a correct knowledge of cover distribution. The main goals for the design of effective steganographic algorithms [12] and [13] are, either to modify the cover as little as possible, or to modify the cover data in inconspicuous parts.

For detecting the presence of hidden messages in multimedia Steganalysis is more commonly used technique. Steganalysis can be applied to digital images and to digital video as reported in [14][15] respectively. Here we are trying to improve steganographic techniques by considering different cover medias to provide high security to the secrete data while transmitting it over the networks. Research paper [16] explains an improved data hiding technique based on BCH coding. The proposed embedder hides data into a block of input data by modifying some coefficients in the block in order to null the syndrome. The proposed embedder can hide data with less computational time and less storage capacity compared to the existing methods. The complexity of the proposed method is linear while that of other methods are exponential for any block size . Thus, it is easy to extend this method to a large. The BCH syndrome coding for steganography is now viable ascribed to the reduced complexity and its simplicity of the proposed embedder.

Research paper [17] proposes two data hiding approaches using compressed MPEG video. The first approach hides message bits by modulating the quantization scale of a constant bit rate video. A payload of one message bit per macroblock is achieved. A second order multivariate regression is used to find an association between macroblock-level feature variables and the values of a hidden message bit. The regression model is then used by the decoder to predict the values of the hidden message bits with very high prediction accuracy. The second approach uses the flexible macroblock ordering feature of H.264/AVC to hide message bits. Macroblocks are assigned to arbitrary slice groups according to the content of the message bits to be hidden. A maximum payload of three message bits per macroblock is achieved. The proposed solutions are analyzed in terms of message extraction accuracy, message payload, excessive bit rate and quality distortion.

II PREVIOUS WORKS

The least-significant-bit (LSB)-based approach is a popular type of steganographic algorithms in the spatial domain. However, we find that in most existing approaches, the choice of embedding positions within a cover image mainly depends on a pseudorandom number generator without considering the relationship between the image content itself and the size of the secret message. Thus the smooth/flat regions in the cover images will inevitably be contaminated after data hiding even at a low embedding rate, and this will lead to poor visual quality and low security based on our analysis and extensive experiments, especially for those images with many smooth regions.

An efficient approach has been proposed [20] by expanding the LSB matching revisited image steganography and proposed an edge adaptive scheme which can select the embedding regions according to the size of secret message and the difference between two consecutive pixels in the cover image. For lower embedding rates, only sharper edge regions are used while keeping the other smoother regions as they are. When the embedding rate increases, more edge regions can be released adaptively for data hiding by adjusting just a few parameters. The experimental results evaluated on 6000 natural images with three specific and four universal steganalytic algorithms show that the new scheme can enhance the security significantly compared with typical LSB-based approaches as well as their edge adaptive ones, such as pixel-value-differencing-based approaches, while preserving higher visual quality of stego images at the same time.

Picture quality and statistical undetectability are two key issues related to steganography techniques. The research paper [21] proposed closed-loop computing framework that iteratively searches proper modifications of pixels/coefficients to enhance a base steganographic scheme with optimized picture quality

and higher anti-steganalysis capability. To achieve this goal, an anti-steganalysis tester and an embedding controller—based on the simulated annealing (SA) algorithm with a proper cost function—are incorporated into the processing loop to conduct the convergence of searches. The cost function integrates several performance indices, namely, the mean square error, the human visual system (HVS) deviation, and the differences in statistical features, and guides a proper direction of searches during SA optimization.

III METHODOLOGY

MPEG Video Files as cover media to hide secrete data.

In this work we proposes a novel technique for data hiding in which message bits are hidden by modifying the quantization scale of MPEG video coded with constant bit rates. Here we use the term macroblock. Macroblock is the processing unit in image and video compression formats based on linear block transforms, such as the discrete cosine transform. A macroblock typically consists of 16 X 16 samples, and further subdivided into transform blocks, and may further subdivided into prediction blocks. To extract future variables individual macroblocks are used and second order regression model is computed. To predict the content of the hidden message based on macroblocks level features, the decoder uses regression model.

Hide the message using Quantization Scale Modulation(QSM).

By using QSM, to hide a message, the message is first converted into a binary stream of bits. During the MPEG encoding of individual macroblocks, the message bits are read one at a time. For each coded macroblock, the quantization scale is either incremented or decremented based on the corresponding message bits. The algorithm for hiding message bits in one macroblock is shown in Fig 1. To extract the message from the bit stream, macroblock-level feature variables are extracted during the encoding process. Once the whole message is hidden, we will get feature matrix as well as a message vector. We will then treat the feature matrix as predictors, and the message bits as a response variable and use multivariate regression to compute a prediction model. Once computed, the prediction model can be used to predict the message bit hidden in a given macroblock based on its feature variables. Figure 2 and Figure 3 shows the simple algorithm for message hiding at the encoder and message prediction at the decoder respectively.

Computation of Macroblock level feature variables:

The following feature variables are computed from a MPEG-2 video stream for each coded macroblock. The first feature is from uniform distribution model. The value of virtual buffer discrepancy is calculated. This virtual buffer discrepancy is computed by

$$d^t_j = d^t_{0+Bj-1} - (T_t * (j-1) / \#MBs) \quad (1)$$

where the subscript indicates a macroblock $\#MBs$ index, indicates the total number of macroblocks in a video frame and indicates the frame type; I, P, or B. d^t_0 is the initial buffer fullness at the beginning of coding a frame. It is calculated as the accumulated differences between the actual number of coded frame bits minus the target number of frame bits.d_0^t is updated after the encoding of each video frame. Additionally, b_{j-1} indicates the number of bits spent on coding the previous macroblocks in the current frame. Lastly, T_t indicates the target number of bits in the current Group of Pictures (GoP). The overall bit rate and frame rates are important factors here for computation. It also depends on number of bits used for coding the previous frames in the same GoP, the remaining number of P and B frames in the current GoP and the average quantization scale of the previous frames in the same GoP.

1. Start.
2. Compute the quantization scale for each macroblock MB_i ($mquantization_i$).
3. Read the each message bits ($mbit_i$)
4. if ($mbit_j == 1$)
5. if ($mquantization_i < maximum\_mquantization$)


```
6.              Increment mquantization_i.
7.            Compute for each macroblock MB_i
8.            Virtual Buffer Discrepancy d^t_j.
9.            Spatial Activity sactivityj .
10.                 Actual quantization scale of current macroblock MB_i.
11.         else
12.                 Go to line number 24
13.       end if
14. else if (mquantization_i  > minimum_qunatization)
15.           Decrement (mquantization_i).
16.           Compute for each macroblock MB_i
17.                 Virtual Buffer Discrepancy d^t_j.
18.                 Spatial Activity sactivityj  .
19.                 Actual quantization scale of current macroblock MB_i.
20.       else
21.                 Go to line number 24
22.       end if
23.  end if
24.  End
```

Fig 1: Message insertion algorithm for one macroblock.

It shows that at the decoder for each macroblock virtual buffer discrepancy from uniform distribution model can recalculated. Here part of the sequence header contains video bit rate, the frame rate, horizontal and vertical image size. To find out the value of virtual buffer discrepancy, the decoder can use sequence header information and keep track of the number of bits spent on previous frames and previous macroblocks, provided GoP structure is known. If the GoP structure is unknown, the bit stream can be scanned ahead of computing the virtual buffer discrepancy to figure out the total number of P and B frame in a GoP.

The second feature is the spatial activity of the underlying macroblock. This factor is computed from the four non coded luminance blocks of the current macroblock.It is computed using

$$act_j = 1 + min\,(V_{b1}, V_{b2}, V_{b3}, V_{b4}) \quad \text{---------- (2)}$$

where subscript j represents a macroblock index. In frame based coding, the spatial variance of each luminance block is represented as V_{b1}, V_{b2}, V_{b3},V_{b4}. To adoptively modify the value of the quantization scale according to the spatial activity of the current macroblock, the encoder uses this spatial activity. However, since the variance is calculated using the pixel values of the original frame, as opposed to the reconstructed frame, this spatial activity measure is estimated at the decoder using calculation based on reconstructed frame instead.

The actual quantization scale of the current macroblock is the third feature. Macroblock header in video bit stream contains this actual quantization scale.

Message prediction

The second order multivariate regression model is used to formulate the message prediction problem. The response variable in this case is the message binary bits denoted by the vector **m.** Each macroblock has three feature variables and represented as feature matrix **X.** To perform non linear mapping between the predictors or the feature matrix X and the response variable **m** , the dimensionality of the rows or the feature vectors in matrix X is expanded into *rth* order. One approach to expand dimensionality is the reduced model polynomial expansion[18].

$$X = \begin{bmatrix} x_{1,1} & x_{2,1} & x_{3,1} \\ \vdots & \vdots & \vdots \\ x_{1,n} & x_{2,n} & x_{3,n} \end{bmatrix}.$$

Feature Matrix

To arrange the predictors or the feature vectors of n macroblocks feature matrix is used. This matrix is denoted X which is shown above. The subscripts of the matrix elements indicate the index of feature variables and the number of macroblocks, respectively.

Message Extraction

To extract the hidden message from a coded video, the feature variables of each macroblock are computed and/or extracted from the bitstream. The feature vectors are consequently arranged into a feature matrix and expanded to the second order, resulting in matrix P. The feature matrix is multiplied by the model weights _^{opt} to generate the predicted hidden message M'.

$$m' = P * \_^{opt.} \quad (3)$$

1. Start
2. Input the master file MPEG video with n MBs.
3. Extract the *macroblock feature vectors*.
4. Calculate *macroblock feature vectors with n x 3 feature matrix.*
5. Apply second order polynomial expansion to get reduced model.
6. Calculate *n x 12 feature matrix to get the Model Training.*
7. Input the secrete message bits
8. Embed secrete message bits into the encoder.

Figure 2-Algorithm for Message hiding at the encoder.

1. Start
2. Input the master file MPEG video with n MBs.
3. Extract the *macroblock feature vectors*.
4. Calculate *macroblock feature vectors with n x 3 feature matrix.*
5. Apply second order polynomial expansion to get reduced model.
6. Calculate *n x 12 feature matrix to get the Model Training.*
7. Input the secrete message bits
8. Embed secrete message bits into the encoder.

Figure 3-Algorithm for message prediction at the decoder

Feature	MB_i	MB_{i+1}
Buffer Occupancy	8.6632	0.4338
MB Spatial Activity	0.6765	8.5764
Quantization Scale	9	5

Table I-Example Macroblock Feature Variables

Model Weights	Feature Vector of MB_i	Feature Vector of MB_{i+1}
0.745	1	1
-216.1240	0.7143	0.5432
-216.2430	9.86354	9.82768
-219.1235	9	5
-0.3090	79	23
0.3010	97.08765	96.8907
-183.8910	172.60	76.2345

Table II- Examples Model Weights and Expanded Feature Vectors

The above Table II shows an example of Model weights and the expanded feature vectors of MB_i and MB_{i+1}.To decode the message, the decoder computes the feature variables of the MBs from the encoded bitstream, expands them to the second order and uses the model weights to predict the message bits. The Table I shows an example of Macroblock Feature Variables.

IV EXPERIMENTAL RESULTS

This section reports experimental result of the message hiding techniques and compares them with existing report [22] [23]. For better comparison, we referred the same video sequence which is shown in table III. Prediction accuracy is computed by decoding the video sequence, extracting macroblock based features and arranging them into feature matrix. Proposed work shows that accuracy of message prediction is higher than the existing work [22][23]. The table IV shows that accuracy of message prediction in second order regression is higher than the first order. It also shown in the table that some of the sequence are having very high accuracy. These indicate that proposed work can be used for hiding larger data and achieve very high accuracy in message prediction.

Table III-Video Test Sequences

Sequence ID	Sequence Name	#MBs/ frame	Frames /sec
V1	Coastguard	385	30
V2	Container	385	30
V3	Flower Garden	327	30
V4	Foreman	385	30
V5	Hall Monitor	385	30
V6	Mobile	385	30

Table IV-Accuracy of Message Prediction.

Sequence Name	Poly.Expansion	
	1st Order	2nd Order
Coastguard	93.2%	97.8%
Container	97.7%	99.87%
Flowergarden	77.5%	90.03%
Foreman	91.3%	96.89%
Hall Monitor	96.7%	98.89%
Mobile	73.4%	87%
Average	**88.3%**	**95.08%**

RMSE between Spatial activities of Original and reconstructed macroblocks.

Sequence	RMSE
Coastguard	55.1
Container	32.0
Flower garden	212.8
Foreman	31.2
Hall monitor	33.1
Mobile	178.8

Quantization Scale Method Experimental Results

We analyze the quantization scale message hiding solution using the following criteria.

1. Message prediction accuracy.
2. Message hiding payload which can be measured in kilobits per second (Kb/s)
3. The excessive bit rate as a result of message hiding in Kb/s.
4. Drop in PSNR measured in decibels. The prediction accuracy is computed by decoding a video sequence, extracting macroblock features, and arranging them into feature matrix

Table V-Review [19]

Sequence Name	Payload Kbits/s	Bitrate Overhead Kbits/s	Average distortion in dB
Coastguard	2.70	284.02	0
Container	0.55	83.08	0
Flower garden	3.57	502.32	0
Foreman	2.27	206.26	0
Hall Monitor	1.23	109.49	0
Mobile	4.36	464.30	0
Average	**2.45**	**274.91**	**0**

Table VI-Proposed Solution

Sequence Name	Payload Kbits/s	Bitrate Overhead Kbits/s	Average distortion in dB
Coastguard	11.27	0	0.24
Container	11.29	0	0.85
Flower garden	9.10	0	0.10
Foreman	11.65	0	0.39
Hall Monitor	11.97	0	0.62
Mobile	10.42	0	0.10
Average	**10.95**	**0**	**0.38**

Existing works in terms of Payload, Overhead and Distortion is illustrated in Table V. Proposed work in terms of Payload, Overhead and Distortion is illustrated in Table VI. Both existing and proposed solutions are compared.

V CONCLUSION

We studied one of the novel approach for message hiding. In this approach, the quantization scale of a CBR video is either incremented or decremented according to the underlying message bit. A second-order multivariate regression is used to associate macroblock-level features with the hidden message bit. The decoder makes use of this regression model to predict the message bits. It was shown that high message prediction accuracy that is consistently nearing 97% - 99% can be achieved. Compared to other existing works our proposed solution is better in terms of accuracy of message prediction. However, the message payload is restricted to one bit per macroblock. Comparisons with other existing work revealed the effectiveness of the solutions in terms of message payload, video distortion and excessive overhead. Here it indicates that bit rate overhead is almost zero but increases small amount of payload and distortion. Future work can be of achieving negligible amount of distortion with bit rate over head is zero. And also we studied number of research papers related to data transmission techniques over the networks using steganography.

REFERENCES

[1] S. Kapotas and A. Skodras, "A new data hiding scheme for scene change detection in H.264 encoded video sequences," in *Proc. IEEEInt. Conf.Multimedia Expo ICME*, Jun. 2008, pp. 277–280.

[2] M. Carli, M. Farais, E. D. Gelasca, R. Tedesco, and A. Neri, "Quality assessment using data hiding on perceptually important areas," in *Proc.IEEE Int. Conf. Image Processing, ICIP*, Sep. 2005, pp. III-1200-3–III- 1200-3.

[3] Y. Li, H.-X. Chen, and Y. Zhao, "A new method of data hiding based on H.264 encoded video sequences,"in *Proc. IEEE Int. Conf. Signal Processing, ICSP*, Oct. 2010, pp. 1833–1836.

[4]K. Nakajima, K. Tanaka, T. Matsuoka, and Y. Nakajima, "Rewritable data embedding on MPEG coded data domain," in *Proc. IEEE Int.Conf. Multimedia and Expo, ICME*, Jul. 2005, pp. 682–685.

[5] D.-Y. Fang and L.-W.Chang, "Data hiding for digital video with phase of motion vector," in *Proc. IEEE Int. Symp. Circuits Systems, ISCAS*, Sep. 2006.

[6] C. Xu, X. Ping, and T. Zhang, "Steganography in compressed video stream," in *Proc. Int. Conf. Innovative Computing, Information andControl, ICICIC'06*, 2006, vol. II, pp. 803–806

[7] D.-Y. Fang and L.-W.Chang, "Data hiding for digital video with phase of motion vector," in *Proc. IEEE Int. Symp. Circuits Systems, ISCAS*, Sep. 2006.

[8] Y. Hu, C. Zhang, and Y. Su, "Information hiding based on intra prediction modes H.264/AVC," in *Proc. IEEE Int. Conf. Multimedia and Expo, ICME*, Jul. 2007, pp. 1231–1234.

[9] Rongyue Zhang, VasiliySachnev, Magnus Bakke Botnan, HyoungJoong Kim," An efficient Embedder for BCH Coding for Steganography", IEEE Transactions on Information Theory, Vol. 58, no. 12, December 2012.

[10] Tamer 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, April 2012.

[11] Y. Wang and P. Moulin, "Perfectly secure steganography: Capacity, error exponents, and code constructions," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2706–2722, Jun. 2008.

[12] C. Fontaine and F. Galand, "How Reed–Solomon codes can improve steganographic schemes," *EURASIP JInf. Security*, vol. 2009, pp. 274845-1–274845-10, 2009.

[13] D. Schönfeld and A. Winkler, "Embedding with syndrome coding based on BCH codes," in *Proc. 8th ACM Workshop Multimedia Security*, 2006, pp. 214–223.

[14] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE Trans. Inform. Forensics Security*, vol. 1, no. 1, pp. 111–119, Mar. 2006.

[15] U. Budhia, D. Kundur, and T. Zourntos, "Digital video steganalysis exploiting statistical visibility in the temporal domain," *IEEE Trans.Inform. Forensics Security*, vol. 1, no. 4, pp. 502–516, Dec. 2006.

[16] An Efficient Embedder for BCH Coding for Steganography ,Rongyue Zhang, VasiliySachnev, Magnus Bakke Botnan, HyoungJoong Kim*, Member, IEEE*, and Jun Heo*, Member, IEEE* .

[17] Data Hiding in MPEG Video Files Using Multivariate Regression and Flexible Macroblock Ordering, Tamer Shanableh, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL.7, NO. 2, APRIL 2012.

[18] K.-A. Toh, Q.-L. Tran, and D. Srinivasan,"Benchmarking a reduced multivariate polynomial pattern classifier,"*IEEETrans.Pattern Anal Machine Intell.*,vol.26,no.6,Jun.2004

[19] K. Wong, K. Tanaka, K. Takagi, and Y. Nakajima, "Complete video quality-preserving data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 10, Oct. 2009.

[20] WeiqiLuo*, Member, IEEE*, Fangjun Huang*, Member, IEEE*, and Jiwu Huang*, Senior Member, IEEE,* "IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 5, NO. 2, JUNE 2010

[21] Guo-Shiang Lin, Yi-Ting Chang, and Wen-Nung Lie, " A Framework of Enhancing Image SteganographyWith Picture Quality Optimization and Anti-Steganalysis Based on Simulated Annealing Algorithm", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 12, NO. 5, AUGUST 2010.

[22] H. A. Aly, "Data hiding in motion vectors of compressed video based on their associated prediction error," *IEEE Trns. Inform. Forensics security*, vol 6. No. 1 pp-14-18, Mar 2011.

[23] K.Wong, K.Tanaka, K.Takagi, and Y.Nakujima, "Complete video quality preserving data hiding," *IEEE Trns.Circuit Syst. Video Technol*, vol 19, no 10, Oct 2009.

