# Adaptive Crypto-Steganosystem for videos based on Information Content and Visual Perception

Prerana Mukherjee*, Siddharth Srivastava*, Brejesh Lall*, Saisha Asolkar†, Meera Pai†
*Indian Institute of Technology, Delhi, India, †National Institute of Technology, Goa, India
*{eez138300, eez127506, brejesh}@ee.iitd.ac.in, †{asolkar.saisha, meerapai93}@gmail.com

***Abstract*—Steganography is the art of hiding secret data inside a carrier media. Most steganographic techniques suffer from the drawback that they are unable to retain the perceptual quality. Using saliency cues for developing an adaptive steganographic technique can help to alleviate this problem. In this work, a novel perception driven robust crypto-steganographic algorithm is proposed for embedding secure data in videos. The proposed scheme selects the payload regions based on natural scene statistics. To further strengthen the scheme and ensure intractability of secure data, the encrypted secret data is embedded in a random manner using jumbling sequence generator in the frames. We utilize perceptual hashing to evaluate the number of bit insertions that will not compromise the perceptual quality. A comprehensive performance evaluation of the proposed scheme is provided to detail the effectiveness. We demonstrate that the scheme shows a lot of promise in being robust against statistical and saliency based attacks.**

***Keywords*- *Saliency, Steganography, Encryption, Perceptual Quality***

## I. INTRODUCTION

With the ever increasing size of the data and the rapid growth of information exchange over the internet, it is of prime importance to safeguard the data from attack by the illegitimate users. The two important dimensions of information hiding are Cryptography and Steganography. Steganography maintains the structure of the hidden message as opposed to cryptography. Thus, steganography does not draw any undesirable attention from the intruder. Similar to cryptographic systems, steganography can also be done using a private (secret key stegano-system) or a public key (public key stegano-system). It finds wide applications in covert writing[1], digital watermarking, modern printers, intelligence services and forensics.

Nowadays, steganography is used in some form of digital media like text [1], images, audio and video files[2]. For images, the spatial domain methods use bit insertion (LSB)[3], masking and noise manipulation schemes. Frequency domain methods work on the image transforms. In most of these methods the perceptual quality of the image is highly degraded and they also suffer from high computational complexity. In [4], authors proposed a fast embedding technique using enhanced Hidden Markov Model for video steganography. The authors in [5] provide a saliency based ROI selection method which helps in embedding the copyright information into the DCT coefficients which is further embedded in wavelet domain to make it resistant to tampering attacks. In this paper, we propose a saliency based crypto-steganography for embedding the encrypted secret information in the videos. This is motivated by the work in [6] in which the authors employ a saliency based block steganography. In our work, we utilize a stronger saliency map which capture the salient regions quite well. We adaptively distribute the data across all the frames so that it is not concentrated in a single region ensuring the data to remain imperceptible inside an innocuous looking cover image. The distribution of the information is so sparse that it is immuned to statistical and saliency based attacks. The statistical features between the stego and the cover frames is high which results in higher rate of misclassification. Our technique employs natural scene statistics to find the frames in the video which have medium information content. We then find the blocks in those frames which are less salient using saliency map. To further strengthen the technique we encrypt the secret message before embedding it inside the cover image. The payload capacity is even more in case of videos as compared to images. Saliency in images is a measure of distinctive image quality assessment. It pops out the most conspicuous part in the image. The motivation for the use of saliency driven steganographic scheme can be attributed to two main reasons. Firstly, saliency based steganography help to retain the perceptual quality which is not achieved by most of the steganographic techniques. Secondly, embedding information in the less salient regions will assist in camouflaging the changes to be perceived by the attacker during visual inspection. Mostly, the steganographic attacks are based on statistical analysis, saliency based methods can provide a good alternative for maintaining the perceptual content and perform better than classical techniques against attacks. In view of the above discussions, the major contributions of this paper are:

1) Development of a block based saliency driven crypto-steganography scheme in videos for embedding encrypted secure data (image or text data) in an imperceptible and intractable manner.
2) Robustness testing of our scheme is done with different methods. The performance evaluation of the stego videos ensures the perceptual quality to be equivalent to the original content.

Rest of the paper is organized as follows. In Section II, we describe the proposed approach in detail. In Section III, we present the results and discussions. Finally, we conclude the paper in Section IV.

978–1–4673–8564–0/15/$31.00 

## II. METHODOLOGY

In this section, we give a detailed overview of the saliency map based crypto-steganographic technique. In the following subsections, we describe the components of the proposed method (Figure 1) in detail.

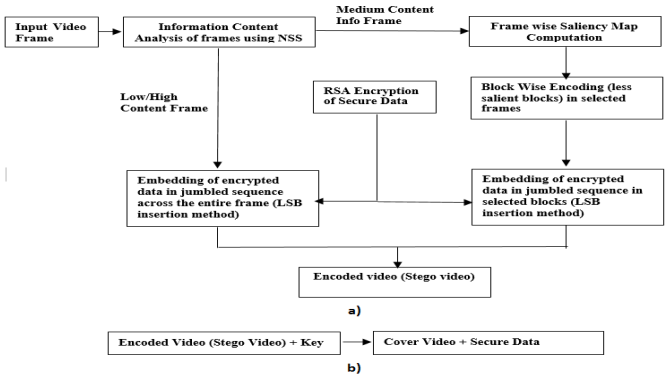



Fig. 1. Proposed Methodology: a) Encoding scheme b) Decoding Scheme

### A. Saliency Computation

Natural scene statistics (NSS) [7] helps in analyzing the relevant information from the statistics obtained from a set of natural images. We use NSS to predict the information content in the frames. In this context, information content means the percentage of salient region in each frame. The value of the information content helps in identifying the medium and high information content frames which are ideal for payload embedding. Natural images generally have low or medium level information. In low information content images, data when embedded in a sequential fashion is more susceptible to steganographic attacks as the changes are easily perceptible. Therefore, in order to get the appropriate ranges for information level, range of information content is divided into bins having intervals as (0.05-0.1,...,0.95-1). We choose a threshold to decide the cut-off ranges for low, medium and high level information. The threshold between low and medium level information is chosen as follows. For each bin the sequential data is embedded into frames till it begins affecting the perceptual content which is measured using Mean Square Error. The bins with majority of perceptually affected frames (more than half) are termed as medium information content bins. In our experiments, we observed that this serves as the optimal range for medium content information. Then a saliency based block-wise steganography of data is performed in all the frames within these medium content bins. As demonstrated in the experiments, this block-wise embedding retains perceptual content against the sequential embedding in the same frames. Range of high content frames was chosen by observation. The frames above a certain bin interval were always found to contain large number of salient regions. The same was verified experimentally on a large number of input frames. This threshold serves as the upper limit for the medium information content frames. The data is embedded in these frames in sequential (for low or high content) and block wise manner (for medium content). In case the video contains both low and medium information frames, the data is first embedded in the blocks of medium information frames and then to the low information frames in sequential manner. This results in adaptive distribution of the data across all the frames with lower perceptual degradation. Figure 2 shows some examples of videos having low, medium and high level information content.

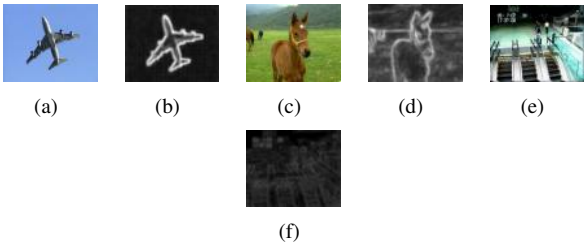



Fig. 2. Information Content Assesment using Natural Scene Statistics. Figure 2(a), 2(c), 2(e): Frame with low (0.1-0.3), medium (0.4-0.7) and high information content (0.8-1 .0) respectively. Figure 2(b), 2(d), 2(f): The corresponding saliency maps for these using NSS.

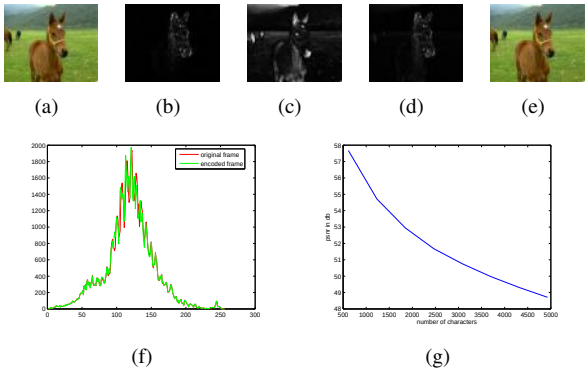
*1) Block-Wise Saliency Map based Steganography :* Saliency is utilized to find the relevant blocks where the data can be safely embedded in the image i.e. without changing the perceptual quality. In our scheme, we utilize the blocks with minimum mean and minimum variance of the corresponding saliency map of the frame which are suitable for payload embedding. This preserves the information content and embeds the data in less salient regions without distorting the perceptual quality. Table I shows that the block based scheme is more efficient than sequential techniques. The global saliency map is computed for the selected frames. The motion map is calculated using Lucas- Kanade optical flow computation. The weighted combination of motion map and the saliency map obtained by the Frequency-Tuned approach [8] give the final saliency map. The maps are assigned weights $w_i = \frac{\sigma_i}{\sum_{i=1}^{n} \sigma_i}$, where $i = 1...n$, $n$ is the number of maps and $\sigma_i$ indicates variance of the $i^{th}$ map. Our proposed block based method helps in identifying the less salient parts for data embedding and decentralization of the data which makes it less prone to attacks. Figure 3 shows the saliency map computation and the encoded frame. In Section II-B, we describe the encryption scheme for secure data embedding.

TABLE I
PERFORMANCE DEGRADATION

| Technique | MSE | PSNR | chi-square similarity |
|---|---|---|---|
| SEQUENTIAL | 0.0119 | 67.3635 | 0.0071 |
| BLOCK-WISE | 0.00004 | 91.8219 | 2.8380e-05 |

### B. Chaotic RSA Encryption

The security of the data in the stego video is further enhanced by embedding the encrypted data in the payload regions. Each character in the secret data is encrypted to fixed length 16 bit binary string using RSA algorithm. To further strengthen the steganographic scheme and ensure intractability of secure data, the encrypted secret data is embedded in a random way using a jumbling sequence generator. This chaotic sequence becomes difficult to trace. In the medium information




(a) (b) (c) (d) (e) (f) (g)

Fig. 3. Figure 3(a): Original Frame (medium content), Figure 3(b): Motion Map, Figure 3(c): Frequency Tuned Map, Figure 3(d): Saliency Map, Figure 3(e): Encoded Frame containing secure data, Figure 3(g): Histograms for original and encoded frame, Figure 3(g): PSNR vs number of characters encoded block-wise in a medium information frame.

frames this scheme is performed in the block-wise manner. In low or high content videos this chaotic RSA encryption is done across all the frames randomly in such a manner that there is less deviation from the original content. Our proposed methodology is explained in Figure 4. To prevent the man-in-the-middle attacks the key should be shared using a separate secure channel.

### C. Perceptual Hashing

Perceptual Hashing (PH) [9] provides a hash value for any multimedia content. We have used PH as a performance measure to verify the perceptual quality and to find the number of bits which can be used for LSB insertion without compromising the data content. First, the perceptual hash value H for the original image is calculated. One bit is modified in every pixel and again hash value H1 is generated. The hamming distance between the hash values H and H1 indicates the maximum change that is induced due to the LSB insertion. Similarly, the process is repeated for every bit change. It is observed that at most 3 bit secure data insertion in every pixel does not bring any noticeable changes in the image. Thus, the data embedding is done as: 3 bits are changed in R channel, 3 bits in G channel and 2 bits in B channel. The number of bits modified for blue channel is attributed to the fact that the cones in the primary visual cortex are more sensitive towards the blue channel.

## III. Experiments and Results

The experimental setup is performed on a 64-bit Windows 8.1 OS with i7-4700MQ processor 2.40 GHz 8 cores, 16 GB RAM system. MATLAB R2014a with Computer Vision Toolbox and Parallel Computing Toolbox was used for implementation. pHash [10] (DCT hash) is used for implementing perceptual hashing.

### A. Dataset

The dataset for videos by Fukuchi et al [11] has been taken for experimentation. This dataset contains 25 videos(uncompressed AVI clips of natural scenes at 12 fps).

**Step 1: RSA algorithm** Two large prime numbers: $p, q$, n= p.q

$$Euler\,Totient\,Function: \phi(n) = (p-1).(q-1) \quad (1)$$

encryption key $e: 1 < e < \phi(n), gcd(e, \phi(n)) = 1$
decryption key d: $e.d = 1 \mod \phi(n), 0 \leq d \leq n$
Public key: $e, n$, Private key: $d, p, q$
Ciphertext $c: m^e \mod n$, Message $m: c^d \mod n$
**Step 2: Jumbling Sequence Generator** Polynomial function: $ax^4 + bx^3 + cx^2 + dx$. Two random numbers are generated to jumble the row and column indices.
For row: $ran\_num = (ax^4 + bx^3 + cx^2 + dx)$ x 100000
For column: $ran\_num1 = (ay^4 + by^3 + cy^2 + dy)$ x 100000
$a, b, c, d$ values are passed as key
$x$, $y$: functions of row and column indices of non-salient blocks

$$x = \frac{j+k}{j.k}, y = |1 - x| \quad (2)$$

for $block(i, j, k)$ where $i$: frame number, $j, k$: pixel coordinates in the block
Jumbling number is obtained as follows: For row: $jumb = |\lceil ran\_num \rceil, bs - 1|$
For column: $jumb1 = |\lceil ran\_num1 \rceil, bs - 1|$
$bs$: block size. Absolute value is taken to ensure that the pixel indices lie within the block.
$\{jumb, jumb1\}$: new pixel indices. For low and high content videos the same method is followed in sequential manner across all frames (not in block-wise manner). Since, we are jumbling the index values it is ensured that same pixel index is not visited twice.
**Step 3: Perceptual Hashing** We calculate the PH to find the number of bit insertion in LSB.
**Step 4: Data Embedding** Take ASCII value of each character in secret text. Encrypt each character using RSA. Encrypted value=16 bit binary number. First 8 bits of the encrypted text is encoded in the LSB bits of RGB channel (R:3 bits, G: 3 bits, B: 2 bits). Remaining 8 bits are embedded in the next pixel (as given by the jumbling sequence generator) in a similar manner. The 16 bit binary string of the encrypted text are splitted as: $a1$: 1 2 3, $b1$: 4 5 6, $c1$: 7 8, $a2$: 9 10 11, $b2$: 12 13 14, $c2$: 15 16. Calculate $a1$: $num = bitwiseAND\{ciphertext, 1110000000000000\}$ i.e. $a1$ is right shifted by 13 positions in $num$. Similarly $b1$, $c1$, $a2$, $b2$, $c2$ are calculated.
Embedding values in pixel: $PV$: pixel value
R channel:
$PV = (PV(RED)\,AND111100)\,OR(a1)$
G channel:
$PV = (PV(GREEN)\,AND111100)\,OR(b1)$
B channel:
$PV = (PV(BLUE)\,AND111100)\,OR(c1)$
This would embed the first 8 bits in encrypted data in the first pixel. Similarly replace $a1$, $b1$, $c1$ by $a2$, $b2$, $c2$ in R, G and B channels of next pixel to encode rest 8 bits of the encrypted data.
**Step 5: Decoding** Reverse Procedure is applied for decoding.

Fig. 4. Algorithm for Proposed Methodology

### B. Perceptual Quality Assessment

The block size is selected as 8x8. The key for encryption consists of the following information: selected frame number; block size; block numbers within the selected frames; a, b, c, d: values for jumbling sequence generator; p, q: RSA Encryption. The secure data which is to be embedded can be text or image. For the experiments, the amount of hidden information in the cover videos consists of 2203 characters which is equivalent to one page of text data. Due to large payload capacity of videos, this amount of data can fit within 2-3 frames of low or medium content videos. So, our proposed scheme is capable of handling even larger amount of data. The performance analysis is done on the basis of the statistical measures: Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Complex Wavelet Structural Similarity (CWSSIM), Feature Similarity (FSIMc) and PH. CWSSIM values lie in the range 0-1. Two images with higher structural similarity will have CWSSIM measure close to 1. PH was calculated with normalized hamming distance. Table II shows the performance measures. Results indicate that the perceptual degradation is negligible thus less susceptible to stego attacks. Figure 5 shows that there is no perceivable difference in the histograms between cover and stego frames indicating that the data is sparsely distributed across frames which is hard to detect by any steganalysis scheme.

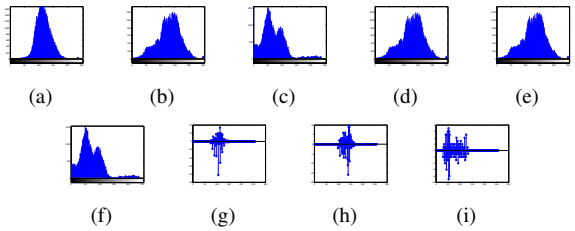



Fig. 5. Figure 5(a),5(b),5(c): Histogram for Red, Green, Blue channel respectively for Cover Frame(medium content), Figure 5(d), 5(e), 5(f): Histogram for R, G, B channel respectively for Stego Frame, Figure 5(g), 5(h), 5(i): Difference between Histograms of Cover and Stego Frames for R, G, B channel.

TABLE II
PERFORMANCE MEASURES ON ORIGINAL VS ENCODED STEGO FRAME.

| Info Content | MSE | PSNR | CWSSIM | PH | FSIMc |
|---|---|---|---|---|---|
| LOW | 0.0498 | 63.0981 | 0.9993 | 0.381944 | 0.9735 |
| MEDIUM | 0.00004 | 91.8219 | 0.9998 | 0.027778 | 0.9939 |
| HIGH | 0.00023 | 84 | 1.0000 | 0.022569 | 0.9992 |

### C. Steganalysis

In our experiments, we applied our crypto-steganographic algorithm on the videos and encoded video sequences were generated. The frames are of the size 288x352. The video size remained unaltered as the cipher text bits just replace the LSBs of the original frame. The dataset consists of the cover frames and their corresponding stego frames. For classification, Support Vector Machine (SVM) with Gaussian kernel is used. The features are obtained from DCT coefficients in the 8X8 blocks of the frames. Table III compares our method against the state-of-the-art steganographic techniques using five-fold cross validation. It can be observed from the results that our technique enforces a majority in the number of misclassified samples i.e cover images which are classified as stego images. This demonstrates that the proposed technique is able to distribute data across the frames effectively and efficiently in addition to minimally altering the perceptual quality of the video. Additionally, we checked the PSNR rates by changing the information embedding rates. For 123 characters/frame, the PSNR values for different methods were: Our method (76), OpenPuff (64) and Steganography using Wavelet transform (79). On changing the number of characters encoded the PSNR values vary to a small extent.

TABLE III
SVM CLASSIFICATION PERFORMANCE USING DCT FEATURES.

| JSteg | StegHide | OutPuff | F5 | OUR METHOD |
|---|---|---|---|---|
| $80.5 \pm 4.2$ | $51 \pm 6.4$ | $65.2 \pm 1.6$ | $78.5 \pm 3.1$ | $49 \pm 7.8$ |

## IV. CONCLUSION

The paper proposes a block based crypto-steganographic scheme with saliency measure being a crucial component for adaptively selecting the blocks for embedding the secure data. The data is distributed according to the level of information content in the frames such that the change after embedding is not perceptible. To further strengthen the scheme, the encrypted text (using RSA) is randomly spread in the blocks. Steganalysis was not able to detect the presence of hidden messages in many stego frames of the encoded video and recognized them as plain cover frames, showing that the proposed scheme is robust against the statistical and saliency based steganalysis techniques.

## REFERENCES

[1] B. Li, J. He, J. Huang, and Y. Q. Shi, "A survey on image steganography and steganalysis," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 142–172, 2011.
[2] M. M. Sadek, A. S. Khalifa, and M. G. Mostafa, "Video steganography: a comprehensive review," *Multimedia Tools and Applications*, pp. 1–32, 2014.
[3] D.-C. Lou and C.-H. Hu, "Lsb steganographic method based on reversible histogram transformation function for resisting statistical steganalysis," *Information Sciences*, vol. 188, pp. 346–358, 2012.
[4] M. Ramalingam and N. A. M. Isa, "Fast retrieval of hidden data using enhanced hidden markov model in video steganography," *Applied Soft Computing*, vol. 34, pp. 744–757, 2015.
[5] L. Tian, N. Zheng, J. Xue, C. Li, and X. Wang, "An integrated visual saliency-based watermarking approach for synchronous image authentication and copyright protection," *Signal Processing: Image Communication*, vol. 26, no. 8, pp. 427–437, 2011.
[6] R. K. Singh and B. Lall, "Saliency map based image steganography," in *Image and Vision Computing New Zealand (IVCNZ), 2013 28th International Conference of*. IEEE, 2013, pp. 430–435.
[7] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "Sun: A bayesian framework for saliency using natural statistics," *Journal of vision*, vol. 8, no. 7, p. 32, 2008.
[8] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*. IEEE, 2009, pp. 1597–1604.
[9] C. Zauner, "Implementation and benchmarking of perceptual image hash functions," *Master's thesis, Upper Austria University of Applied Sciences, Hagenberg Campus*, vol. 43, 2010.
[10] "Phash: the open source perceptual hash library." http://www.phash.org/.
[11] K. Fukuchi, K. Miyazato, A. Kimura, S. Takagi, and J. Yamato, "Saliency-based video segmentation with graph cuts and sequentially updated priors," in *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*. IEEE, 2009, pp. 638–641.