Implementation of Advanced Encryption Standard Algorithm on Steganography

M. Senthil Murugan, T. Sasilatha

Abstract--- It is sometimes not enough to keep the contents of a message secret. It may also be necessary to keep the existence of the message secret. The algorithm that is described here is about hiding an audio file into video frame. A video file is chosen first. The video file is then divided into frames. A particular frame is selected and skin tone detection is performed on it using HSV color space model. The skin region is taken as the Region of Interest (ROI). An audio file is the secret message to be transmitted. It is embedded into the frame using Advanced Encryption Standard (AES) algorithm. The secret audio file encrypted on a particular frame is transmitted with a key. The receiver can extract the secret audio file only if he has the key. This concept is based on visual method. The goal of using a video to encrypt the data is that it will be difficult to detect the frame in which the data is embedded. The secret data can be retrieved only by the user who has knowledge of the frame in which the data is embedded and the key which is known only by the sender and receiver. Thus a secured communication is established.

Keywords--- AES, Steganography, Skin Tone Detection.

I. INTRODUCTION

In the first chapter a video steganography algorithm that targets high imperceptibility as well as PSNR is proposed. Steganography is defined as the art and science of secret communications, primarily concealing the existence of the communication. Steganography works by embedding the secret message in an innocent looking cover which can be transmitted without raising suspicion. Almost any digital file can be used as a steganography cover. Popular cover types include text documents, audio tracks, digital images, videos and DNA-sequences. Data hiding in videos has gained practical significance nowadays due to the huge technological advancement of multimedia systems. Usually, a video have a large number of frames. This paper proposes a blind adaptive video steganography algorithm that targets high imperceptibility as well as robustness to MPEG-4 compression. The proposed algorithm embeds secret images in video files. Human skin regions are selected as ROI (Region of Interest) for the data embedding process. Skin detection is adaptively carried out based on face detection.

1.1 Steganography

In Steganography secret message is the data that the sender wishes to remain confidential. The cover or host is the medium in which the message is embedded and serves to hide the presence of the message. In this paper covers and secret messages are restricted to being digital images. The cover-image with the secret data embedded is called the "Stego-Image". The Stego-Image should resemble the cover image under casual inspection and analysis. In addition, for higher security requirements, we can encrypt the message data before embedding them in the cover-image to provide further protection. For this the encoder usually employs a stego-key which ensures that only recipients who know the corresponding decoding key will be able to extract the message from a stego-image. For proposed method cover image is cropped interactively and that cropped region works as a key at decoding side yielding improved security. While designing the steganographic system, invisibility factor i.e. human eyes should not distinguish the difference between original and stego image should be considered.

1.2 Modern Algorithms

Modern steganography identifies two main classification schemes for the of algorithms. The first distinguishes algorithms based on file type. The second, more widely used scheme categorizes based on embedding method.

1.3 Injection (Insertion) and Substitution

Insertion-based techniques hide data in sections of a file that are ignored by the processing application and do not modify those bits that determine the contents of a file that are relevantto an end-user. For example, an insertion algorithm may write data in the comment blocks of an HTML file. Several file types and programs also establish an EOF (End of File) marker to signify the end of a file. Data written after this marker is nonexistent as far as meaningful content is concerned. However, from a steganography standpoint the EOF can be used to mark the beginning of hidden data. Utilizing an insertion technique changes file size according to the amount of data hidden and therefore can be used to determine the presence of hidden information. Coming across a 5MB HTML file would arouse suspicion, for instance. In a Substitution-based algorithm, the most insignificant bits of information that determine the meaningful content of the original file are replaced with new data in a way that causes the least amount of distortion. Although file size does not change during execution of the algorithm, the amount of data that can be hidden is limited to the amount of insinificant bits in the file. Higher "quality" files (where applicable) tend to contain more bits of insignificant information. The injection and substitution algorithms both require a "covert" file that contains the information to be hidden, and an "overt" file that acts as the host. The generation technique requires only a covert file, as it is used to create the overt file. For instance, the covert file can be used to create a fractal image with unique colors, angles, and line lengths.

Manuscript received January 25, 2019.

M. Senthil Murugan, Research Scholar, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India.
Associate Professor, Department of ECE, St.Joseph's Institute of Technology, Chennai, Tamilnadu. India.
T. Sasilatha, Dean, Academics, AMET (Deemed to by University), Chennai, Tamilnadu. India.

A main flaw of the insertion and substitution techniques is the ability to compare a given file with another instance of the supposedly "same" file. If the file size, MD5 checksum, or anything else is different, it can be assumed that data has been embedded in the file in question. Since the result of a Generation algorithm is an "original" file, the technique is immune to comparison tests.

1.4 Skin Block Map Generation

Skin detection is simply defined as the process of labeling pixels that have skin-color in an image. Skin detection methodologies are mostly biased to the detection of skin pixels based on their color. The reason behind this is that skin color provides robust information against rotations, scaling and partial occlusions. Skin detection plays an important role in a wide range of image processing applications such as face detection, gesture analysis, Content-Based Image Retrieval (CBIR) systems, and recently its application in steganography. In this context, the skin-map is created using the adaptive skin tone detection algorithm. Their algorithm adapts the skin-colour model according to reliably detected faces in a video frame. First, face detection is performed on the frame using Viola Jones face detector. If multiple faces are detected, each face creates a new skin-color model, which allows detection of skin of different colors and under different lighting conditions even within the same frame. They used the YCbCr color model. The resultant map is a binary matrix, where elements corresponding to skin pixels have the value one, while others have the value zero. Additionally, frames with no faces detected are discarded even if they have other skin regions. The Skin-Block-Map (SBM) aims at decreasing the skin-map sensitivity and enhancing the accuracy of the extracted message by pre-processing the cover frame skin-map. This is accomplished by discarding error-prone skin pixels in the skin-map through a blocking strategy. After a skin-map is generated, it is divided into blocks of size b×b. This operation helps identify the blocks that purely consist of skin information. In other words, a block is considered a skin-block if it consists of ones only, otherwise its skin-map pixels are set to zero.

II. LITERATURE SURVEY

2.1 LSB Method for Hiding Secret Data

Shirali-Shahreza et al., (2007) used the LSB (Least Significant Bit) method for hiding a secret message in the frames displayed by the output screens such as electronic advertising billboards. Each displayed frame is considered an image, that is then broken down into small blocks where the secret data is redundantly hidden in a large number of frames. To extract the hidden message, the intended receiver would photograph the displayed frames in order to extract the secret message. Here spatial features of image are used. This is a simplest steganographic technique that embeds the bits of secret message directly into the (LSB) plane of the cover image. In a gray-level image, every pixel consists of 8 bits. The basic concept of LSB substitution is to embed the confidential data at the rightmost bits (bits with the smallest weighting) so that the embedding procedure does not affect the original pixel value greatly.

This work presents a scheme that can transmit large quantities of secret information and provide secure communication between two communication parties. Both steganography and cryptography can be woven into this scheme to make the detection more complicated. Any kind of text data can be employed as secret msg. The secret message employing the concept of steganography is sent over the network .In addition, the proposed procedure is simple and easy to implement. Also, the developed system has many practical, personal and militaristic applications for both point-to-point and point-to-multi-point communications.

2.2 The Use of Texture as Spatial and Temporal Features of the Cover Video

Mansouri J et al., (2009) presented another adaptive method which combined the use of texture as a spatial feature and temporal features of the cover video. Quantized DCT coefficients are extracted from I-frames and each 8 × 8 block is checked against a threshold to decide if it is suitable for hiding (highly textured or include edges). For each qualified block, eight bits of secret data are embedded in eight quantized DCT coefficients that are determined by a secret key. Motion vectors of P and B frames above a certain threshold are utilized for hiding as well. For each qualified motion vector, two secret bits are embedded. The experimental results showed high hiding capacity as it uses both temporal and spatial features of the video stream. The goal of this article is to propose a steganographic method to covert communication as in military application and to increase the capacity while preserving acceptable imperceptibility. Secret data are embedded in a compressed video bitstream adaptively using temporal and spatial features of the video signal with the considera-tion of the human visual system characteristics. In this method, for each I-VOP, secret data are embedded in some AC coefficients of the blocks with high spatial changes. For each P-VOP and B-VOP, secret bits are embedded in horizontal and vertical components of motion vectors with large magnitude, which represent high tempo-ral changes. Experimental results indicate that this algorithm has high visual quality and embedding capacity. Furthermore, the bit rate remains nearly constant. The only drawbacks of this method are: it requires partial de-compression of the cover video and it is restricted to MPEG-4 encoded videos.

2.3 Digital Image Steganography: Survey and Analysis of Current Methods

Cheddad et al., (2009) stated that communicating secret data in an appropriate multimedia carrier, e.g., image, audio, and video files always comes under the assumption that if the feature is visible, the point of attack is evident, thus the goal here is always to conceal the very existence of the embedded data. Steganography has various useful applications.

However, like any other science it can be used for ill intentions. It has been propelled to the forefront of current security techniques by the remarkable growth in computational power, the increase in security awareness by, e.g., individuals, groups, agencies, government and through intellectual pursuit. Steganography's ultimate objectives, which are undetectability, robustness (resistance to various image processing methods and compression) and capacity of the hidden data, are the main factors that separate it from related techniques such as watermarking and cryptography. This paper provides a state-of-the-art review and analysis of the different existing methods of steganography along with some common standards and guidelines drawn from the literature. This paper concludes with some recommendations and advocates for the object-oriented embedding mechanism. Steganalysis, which is the science of attacking steganography, is not the focus of this survey but nonetheless will be briefly discussed.

2.4 Hiding and Detecting Messages in Media Signals

GB Rhoads et al., (2009) stated that the present invention relates generally to audio and video signal processing, e.g., audio or video signal encoding and detection. One claim recites a method of analyzing an embedded media signal to extract a message information signal therefrom. The method recites in combination with other features utilizing a multi-purpose electronic processor configured for: detecting from an embedded media signal embedded code signals corresponding to registration symbols and variable message information symbols, verifying detection of the embedded code signals coffesponding to the registration symbols; and verifying detection of the variable message information symbols by checking correspondence of the values of the variable message information symbols relative to the registration symbols with an expected sequence of values. The multi-purpose electronic processor is configured for analyzing a plurality of fragments of the embedded media signal with greater or lesser confidence in an extracted message information signal being obtained by processing more or less fragments of the embedded media signal, respectively. Of course, other combinations and claims are provided as well.

On embedding signature codes into motion pictures there is a distinction made between the JPEG standards for compressing still images and the MPEG standards for compressed motion images, so two should there be distinctions made between placing invisible signatures into still images and placing signatures into motion images. As with the JPEG/ MPEG distinction, it is not a matter of different foundations, it is the fact that with motion images a new dimension of engineering optimisation opens up by the inclusion of time as a parameter. A section on how MPEG is not merely applying JPEG on a frame by frame basis will be the same with application of the principles of this invention generally speaking the placement of invisible signatures into motion image sequences will not be simply independently placing invisible signatures into one frame after the next.

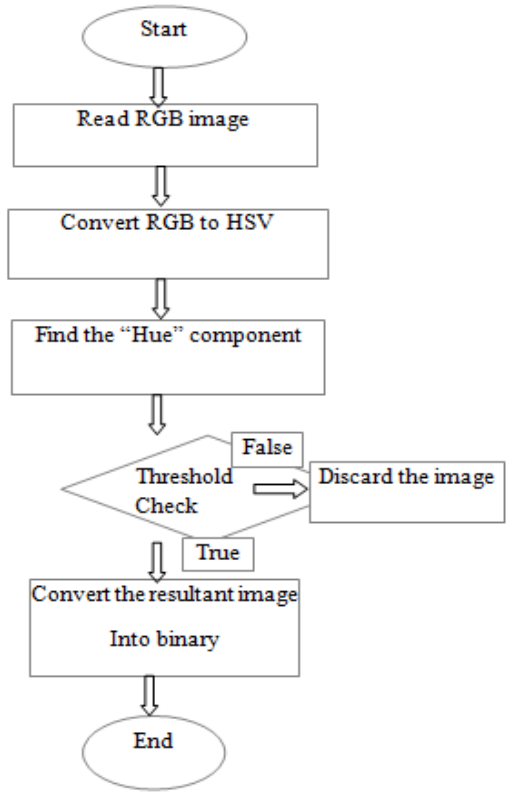
A variety of time based considerations come into play, some dealing with the psychophysics of motion image perception others driven by simple cost engineering considerations.

III. COMPARISION OF THE PROPOSED METHOD WITH OTHER METHODS

The capacity and imperceptibility of the proposed method were evaluated in comparison with four similar approaches. These approaches were selected for comparison as they are designed to be robust methods and they used standard videos for their experimental results namely salesman and car phone video sequences. The table lists the capacity and the calculated PSNR ofthe proposed method and each of the compared methods. The results show that the proposed method outperforms all of the compared techniques in capacity as well as imperceptibility. The only exception is recognized in comparison with Cheddad et al. algorithm, where higher capacity and imperceptibility were achieved. However, the decrease in the proposed method hiding capacity originates from the four factors. That is, first, Cheddad et al. use a rough skin detection algorithm with high false positive detections, while the proposed method use the more precise, Khan et al. skin detection algorithm which has lower false positive detections which affects the total number of skin pixels available for data hiding. Secondly, the proposed method discards frames with no faces detected for achieving precise skin detection. Thirdly, the proposed method adds a skin-map blocking step which sacrifices some skin pixels that may cause errors at extraction for the sake of increasing the robustness. Fourthly, the proposed method uses a wavelet quantization-based data embedding method which applies 3-level Discrete Wavelet Transform (DWT) on the used color channels shrinking their sizes for enhancing the overall robustness. On the other hand, Cheddad et al. algorithm recorded higher PSNR values than the proposed method due to embedding in the Y channel of the YCbCr color space which causes less color distortion. However, both methods achieved PSNR values above 50 dB. Furthermore, raw video files usually have a huge amount of data, so they must be compressed before transmission. This implies robustness of the embedded bits against lossy compression. In this set of experiments, robustness against MPEG-4 compression was evaluated in comparison with Cheddad et al. method. The similarity between the extracted and original messages was calculated after the stego-video is MPEG-4 compressed. The table reports the similarity values measured. The results clearly show that the proposed method outperforms Cheddad et al. Method in robustness against MPEG-4 compression. This was primarily achieved through enhancing both the ROI selection process and the data embedding algorithm. In fact, a more precise selection of the ROI was achieved by using a skin detection algorithm with lower number of false positives and applying a skin-map blocking step for eliminating error-prone skin pixels.

Additionally, a more robust data embedding algorithm was used which applies a multi-level DWT for robustly embedding data by quantizing the relevant detail coefficients. Although, the proposed method succeeded in enhancing the robustness, but this was achieved on the expense of computational complexity and hiding capacity. In fact, the proposed method is more computationally complex as a result of adding the skin-map blocking step and applying three-level DWT on each frame for embedding data. On the other hand, the hiding capacity was decreased to discarding frames with no faces detected, adding the skin-map blocking step and embedding data in some of the third-level DWT detail coefficients.

IV. FLOWCHART FOR SKIN DETECTION



V. PROPOSED SYSTEM

5.1 Frame Separation

The video used is 11 seconds long. The video is converted into 292 frames. The individual frames are stored separately. For encrypting the secret data the frame 100 is selected. In the 100^{th} frame the secret audio file is encrypted.

5.2 Colorspaces Used For Skin Modelling

Colorimetry, computer graphics and video signal transmission standards have given birth to many colorspaces with different properties. A wide variety of them have been applied to the problem of skin color modelling.

5.2.1 RGB

RGB is a colorspace originated from CRT (or similar) display applications, when it was convenient to describe color as a combination of three colored rays (red, green and blue). It is one of the most widely used color spaces for processing and storing of digital image data. However, high correlation between channels, significant perceptual non-uniformity, mixing of chrominance and luminance data make RGB not a very favorable choice for color analysis and color based recognition algorithms.

5.2.2 Normalized RGB

Normalized RGB is a representation, that is easily obtained from the RGB values by a simple normalization procedure: r = R R+G+B g = G R+G+B b = B R+G+B as the sum of the three normalized components is known (r + g + b = 1), the third component does not hold any significant information and can be omitted, reducing the space dimensionality. The remaining components are often called ”pure colors”, for the dependance of r and g on the brightness of the source RGB color is diminished by the normalization. A remarkable property of this representation is that for matte surfaces, while ignoring ambient light, normalized RGB is invariant (under certain assumptions) to changes of surface orientation relatively to the light source.

5.2.3 Skin Detection Using HSV Color Space

The algorithm for the detection of human skin color in color images is explained as follows.

1. Input image is obtained from the Video file which is divided into frames..
2. Input image in RGB color space is converted into HSV color space using transformation. HSV image is a collection of three different images as hue, saturation and value.
3. Masking is applied for skin pixels in the test image.
4. Threshold is applied to the masked image.
5. Threshold image is smoothened and filtered.
6. The output image contains only skin pixels.

VI. AES

AES is a block cipher with a block length of 128 bits. AES allows for three different key lengths: 128, 192, or 256 bits. Most of our discussion will assume that the key length is 128 bits. With regard to using a key length other than 128 bits, the main thing that changes in AES is how you generate the key schedule from the key. Encryption consists of 10 rounds of processing for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for 256-bit keys. Except for the last round in each case, all other rounds are identical. Each round of processing includes one single-byte based substitution step, a row-wise permutation step, a column-wise mixing. The order in which these four steps are executed is different for encryption and decryption. To appreciate the processing steps used in a single round, it is best to think of a 128-bit block as consisting of a 4 ×4 matrix of bytes.

VII. RESULTS

The proposed method produced a PSNR value of 83.50 which is 60% more than the existing method. The similarity of the images is also very high and is 1. The PSNR was found out by calculating the MSE(Mean Square Error) value. The MSE was calculated between the original image and the image at the reciever side. The formula used was:

mse = sum(sum((img_orig(1,1) - I4(1,1)).^2)) / (row * col)

psnr = 10 * log10(255 * 255 / mse)

Fig 7.1: PSNR calculation

The similarity between the original image and the recieved image is calculated using the formula:

ssimval = ssim(I2,I4)

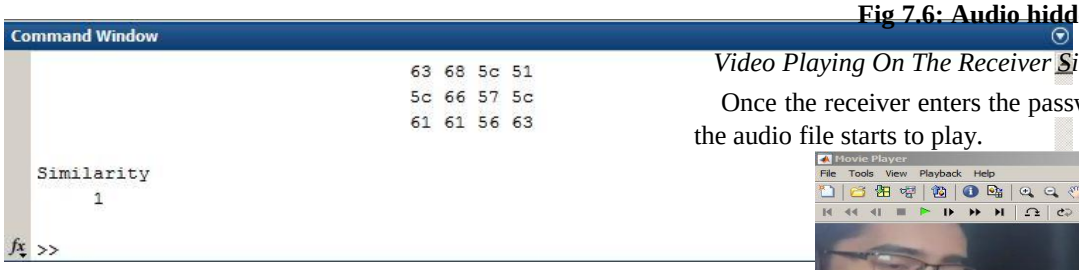


Fig 7.2: Similarity

Original Image

The original image is the frame that is selected out of the 292 frames. The original image is shown below. The image shown below is the 100th frame of the video.

Fig 7.3: Original image

Images After Skin Tone Detection

On the selected frame skin tone detection is performed. A bit map is generated for the image to separate the skin regions from the other portions of the image. The skin region is represented by bit 1 and the other regions are represented by bit 0.

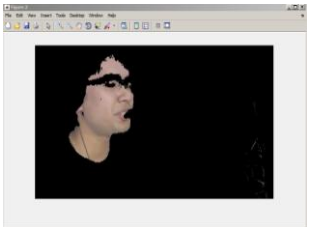


Fig 7.4: Image after skin tone detection

Binarized Image With Skin Block Map

The image is converted to a binary image. On the image after skin tone detection, the skin block map is generated. The audio file is embedded in the skin block region.

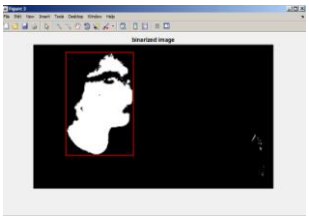


Fig 7.5: Binarized image with skin block map

Audio Hidden Image

The skin block region is alone taken and the audio file is hidden in it. The audio embedded image looks like figure 7.6

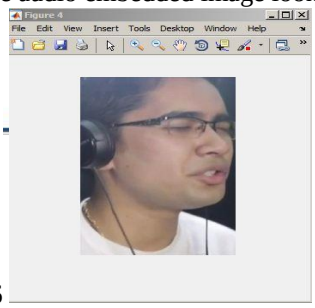


Fig 7.6: Audio hidden image

Video Playing On The Receiver Side

Once the receiver enters the password to start the process, the audio file starts to play.

Fig 7.7: Video playing on the receiver side

Decrypted Image

The audio file is successfully decrypted in the frame 100. The audio file plays only if the same four bit password is given.

Fig 7.8: Decrypted Image

VIII. CONCLUSION AND FUTURE ENHANCEMENT

Providing security for the transmission of the data is one of the greatest challenges in communication. In the proposed method the video is chosen. It is then converted into frames. A 11 second video is chosen. It generated 292 frames. A particular frame is selected among the 292 frames for hiding the secret data. The secret data to be transmitted here is an audio file. Skin tone detection was performed on the selected frame using HSV along with RGB and YCbCr color space models. The image after skin tone detection is taken and it is binarized. Skin block is generated. The audio file is then embedded in the image using AES. A four bit password is provided for security. The audio file is then transmitted. The receiver has to provide the same four bit password to recover the audio file. The PSNR was found to be 83.50 and the similarity was found to be 1.

Future work can be done on increasing the PSNR and similarity by using other encryption algorithms. The throughput can also be increased

REFERENCES

1. Basilio JAM, Torres GA, Sánchez Pérez G, Medina LKT, Meana HMP (2011) Explicit image detection using YCbCr space color model as skin detection. Appl Math Comput Eng.
2. Channalli S, Jadhav A (2009) Steganography an art of hiding data. Int J Comput Sci Eng 1(3):137–141
3. Cheddad A, Condell J, Curran K, Mc Kevitt P (2009) A skin tone detection algorithm for an adaptive approach to steganography. Signal Process 89(12):2465–2478
4. Chen WY (2007) Color image steganography scheme using set partitioning in hierarchical trees coding, digital Fourier transform and adaptive phase modulation. Appl Math Comput 185(1):432–448
5. Elgammal A, Muang C, Hu D (2009) Skin detection - a short tutorial. In Encyclopedia of Biometrics. 1218–1224. Springer US.
6. Eltahir ME, Kiah LM, Zaidan BB, Zaidan AA (2009) High rate video streaming steganography. In: International Conference on Future Computer and Communication (ICFCC 2009) 672–675
7. Farag H, El-Khamy SE (2014) Blind key steganography based on multilevel wavelet and CSF. Int Refereed J Eng Sci 3.
8. Fleck MM, Forsyth DA, Bregler C (1996) Finding naked people. In Computer Vision – ECCV'96: 593–602
9. Gong X, Lu HM (2008) Towards fast and robust watermarking scheme for H.264 video. In Proc. 10th IEEE ISM: 649–653
10. Hamad SH, Khalifa AS (2013) A quantization-based image watermarking using multi-resolution wavelet decomposition. Egypt Comput Sci J.
11. Hu S, KinTak U (2011) A novel video steganography based on non uniform rectangular partition. In: IEEE 14th International Conference on Computational Science and Engineering (CSE) 57–61
12. Kakumanu P, Makrogiannis S, Bourbakis N (2007) A survey of skin-color modeling and detection methods. Pattern Recogn 40(3):1106–1122 Multimed Tools Appl.